



# Modification des voies de repliement d'une petite protéine riche en ponts disulfure : la toxine alpha de *Naja nigricollis*

Grégori Gross

## ► To cite this version:

Grégori Gross. Modification des voies de repliement d'une petite protéine riche en ponts disulfure : la toxine alpha de *Naja nigricollis*. Sciences du Vivant [q-bio]. Museum national d'histoire naturelle - MNHN PARIS, 2008. Français. NNT : . tel-00364212

**HAL Id: tel-00364212**

**<https://theses.hal.science/tel-00364212>**

Submitted on 25 Feb 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**MUSEUM NATIONAL**



**D'HISTOIRE NATURELLE**

**Ecole Doctorale Sciences de la Nature et de l'Homme – ED 227**

**Année 2008**

**N°attribué par la bibliothèque**

|||||

## **THESE**

**Pour obtenir le grade de**

**DOCTEUR DU MUSEUM NATIONAL D'HISTOIRE NATURELLE**

**Spécialité : Biochimie**

**Présentée et soutenue publiquement par**

**Grégori GROSS**

**Le 20 octobre 2008**

---

# **Modification des voies de repliement d'une petite protéine riche en ponts disulfure :**

**La toxine alpha de Naja nigricollis**

---

Sous la direction de : **Monsieur Daniel GILLET, Professeur de l'INSTN**

Et à titre posthume de : **Monsieur André MENEZ, Président du museum national  
d'histoire naturelle**

### **Composition du jury :**

**M. Phillippe BILLIALD**

**Professeur des université, Muséum national d'histoire naturelle, Paris (075)**

**Président**

**M. Daniel GILLET**

**Professeur de l'INSTN, CEA, Saclay (091)**

**Directeur de Thèse**

**M. Alain CHAFFOTTE**

**Chargé de recherche, Institut Pasteur, Paris (075)**

**Rapporteur**

**M. Jacques CHOMILIER**

**Directeur de recherche, Université Pierre et Marie CURIE, Paris (075)**

**Rapporteur**

**M. Pascal DREVET**

**Chargé de recherche, CEA, Saclay (091)**

**Examineur**





*à Paul, mon fils, né le mois dernier,*

*à André Menez qui nous a quitté au début de cette année.*



Je remercie tout d'abord **Philippe Billiald**, **Alain Chaffotte** et **Jacques Chomilier** d'avoir accepté de juger ce travail.

Je souhaite remercier à titre posthume la personne qui a été l'initiateur de ce projet et qui a eu confiance en moi pour que j'y participe : **André Menez**. Je remercie également **Daniel Gillet** d'avoir accepté de prendre la direction de ma thèse à la suite du décès d'André Menez.

Je remercie évidemment les deux autres membres de l'équipe qui a réalisé ce travail : **Pascal Drevet** qui m'a encadré pendant ces trois années de thèse et qui m'a formé à la biochimie des protéines ; **Marie Vandamme** pour l'abnégation dont elle a fait preuve dans sa participation à ce travail.

Je suis également très reconnaissant envers les autres personnes qui ont pris part directement à ce projet à savoir :

**Sophie Zinn-Justin**, **Matthieu Gallopin** et **Bernard Gilquin** (LBSR, CEA-Saclay) qui m'ont encadré et ont participé à l'étude de R.M.N,  
**Enrico Stura** et **Renée Menez** pour l'étude cristallographique,  
**Andrej Galat** (SIMOPRO, CEA-Saclay) pour son travail sur l'analyse des séquences et des structures des protéines à trois doigts,  
**Laurent Bellanger** (CEA –Marcoules) qui a produit les anticorps,  
**Jean Pierre Le Caer** (ICSN, CNRS-Gif/Yvette) pour l'analyse de masse des intermédiaires,  
**Loïc Martin** et **Grégoire Martin** pour m'avoir formé à l'utilisation du Biacore,  
**Marie Anne Lelait** (SIMOPRO, CEA-Saclay) qui a produit la toxine tritiée,  
**Alain Lecoq** (SIMOPRO, CEA-Saclay) qui a produit les peptides épitopes,  
**Fabrice Beau** (SIMOPRO, CEA-Saclay) pour l'acquisition d'images au  $\beta$ -imager,  
**Denis Servent** (SIMOPRO, CEA-Saclay) pour m'avoir formé aux tests d'activité biologique.

Je remercie aussi tout le personnel des deux services qui ont accueilli notre équipe pendant ces trois années à savoir le SIMOPRO (ex DIEP) et le SBIGeM du CEA Saclay.

Je remercie la direction de l'IBiTec-S ainsi que **Michel Werner** de m'avoir offert les moyens de terminer ma thèse dans de bonnes conditions.

Je souhaite aussi remercier les personnes qui m'ont conseillé et permis de m'orienter vers la recherche, à savoir **Laurent Laloux** de l'AFSSA (Maisons Alfort), **Jean Michel Camadro** (Institut Jacques Monod) et **Bertrand Friguet** (Université Paris 7).

Enfin, un grand merci à ma famille :

à mes parents qui m'ont transmis la curiosité scientifique,  
à ma sœur, mon frère et mon beau-frère,  
à mes beaux-parents qui m'ont beaucoup aidé à me décontracter les week-ends,

à mon épouse Claire, qui, en soutenant mon projet d'orientation vers la recherche, m'a aidé à oser réaliser mes rêves,  
à mon fils Paul qui est né le 28 juin 2008 et qui a eu la gentillesse de ne pas trop pleurer la nuit pendant la rédaction de ce manuscrit.



INTRODUCTION	11
1 <i>Le repliement des protéines : un défi majeur de la recherche en biologie</i>	13
1.1 De l'intérêt d'étudier le repliement des protéines	13
1.2 L'expérience originelle	14
1.3 Théories décrivant le repliement des protéines	14
1.4 Méthodes expérimentales de caractérisation du chemin de repliement des protéines sans ...	17
2 <i>Le repliement cotraductionnel des protéines</i>	18
2.1 Arguments contre	18
2.2 Arguments pour	18
2.3 Simulation du repliement cotraductionnel	21
3 <i>Le repliement oxydant in vitro des protéines à ponts disulfure</i>	23
3.1 La réaction élémentaire de formation d'un pont disulfure	23
3.2 Les expériences de repliement oxydant	26
3.3 Classification des mécanismes de repliement des protéines à ponts disulfure	28
3.4 Structures d'intermédiaires de repliement ou de leurs analogues	31
4 <i>Le modèle d'étude : la toxine <math>\alpha</math> de <i>Naja nigricollis</i>, une protéine à trois doigts</i>	34
4.1 La toxine $\alpha$	34
4.2 Fonctions exercées par les toxines à trois doigts	36
5 <i>La famille structurale des protéines à trois doigts</i>	38
5.1 Le domaine à trois doigts	38
5.2 Classification du domaine à 3 doigts au sein du groupe des petites protéines riches en ponts...	38
5.3 Déterminants structuraux conservés au sein de la famille des protéines trois doigts	41
6 <i>Le repliement des toxines à trois doigts</i>	47
6.1 Repliement des neurotoxines trois doigts	47
6.2 Repliement de $\alpha 62$	48
7 <i>Problématiques</i>	50
7.1 Mécanisme de repliement in vitro de la toxine $\alpha$ de <i>N. nigricollis</i>	50
7.2 Vectorisation in vitro du repliement de la toxine $\alpha$ de <i>N. nigricollis</i>	50
MATERIELS ET METHODES	53
1 <i>Production par voie recombinante des mutants étudiés</i>	55
1.1 Séquences et gènes des protéines étudiées	55
1.2 Production des gènes	55
1.3 Clonage amplification et vérification du gène	56
1.4 Expression et purification	58
2 <i>Marquage des toxines</i>	61
2.1 Marquage à la biotine	61
2.2 Marquage à l'azote 15	61
2.3 Marquage au tritium	61
2.4 Marquage au soufre 35	61
3 <i>Repliement oxydant in vitro de la toxine <math>\alpha</math></i>	62
3.1 Réaction de repliement	62
3.2 Analyse des mélanges réactionnels	62
3.3 Modélisation des cinétiques	63
3.4 Analyse des formes intermédiaires	64
4 <i>Production et caractérisation des anticorps</i>	67
4.1 Production des anticorps	67
4.2 Purification et concentration	68
4.3 Caractérisation des anticorps	68
5 <i>Effet des anticorps sur le repliement</i>	76
5.1 Expérience de repliement suivi par mesure de l'activité biologique	76
5.2 Expérience de repliement suivi en isoélectrophorèse	78

RESULTATS	79
1 <i>Le mécanisme de repliement in vitro de la toxine <math>\alpha</math></i>	82
1.1     La longueur de lk1 influe sur les étapes finales du repliement	82
1.2     La longueur de lk1 détermine la productivité des intermédiaires à trois ponts	91
1.3     Interprétation structurale des différences de mécanisme de repliement	101
2 <i>Vectorisation du repliement d'une protéine in vitro</i>	113
2.1     Production de solutions polyclonales d'anticorps	113
2.2     Caractérisation des anticorps obtenus	120
2.3     Effet des anticorps sur le repliement de la toxine $\alpha$	130
2.4     Conclusions concernant la vectorisation	139
DISCUSSION	141
1 <i>Le repliement oxydant de la toxine <math>\alpha</math> et l'influence de lk1</i>	143
1.1     La vitesse de repliement des neurotoxines courtes est dépendante de la longueur de lk1	143
1.2     Infirmerie de l'hypothèse d'une augmentation du nombre d'intermédiaires	143
1.3     Le comportement cinétique de deux formes à trois ponts détermine l'efficacité globale	144
1.4     Un basculement de voie productive modifie la cinétique	145
1.5     Explication structurale de la différence de productivité de la voie C	146
1.6     Origine de la stabilisation de la forme structurée de l'intermédiaire C	148
1.7     Interprétation structurale de la perte d'efficacité de la voie D	149
1.8     Généralisation des résultats à la famille des neurotoxines courtes	151
2 <i>Vectorisation in vitro du repliement de la toxine <math>\alpha</math></i>	152
2.1     Influence de la vectorisation sur le repliement	152
2.2     Principe de la vectorisation du repliement par des anticorps	153
2.3     Les anticorps obtenus	153
2.4     Un seul anticorps répond au cahier des charges	154
2.5     Vectorisation du repliement de protéines multidomaines	158
BIBLIOGRAPHIE	159
ANNEXES	167

$^1\text{H}$   $^1\text{H}$  COSY : Proton proton correlation spectroscopy  
 $^1\text{H}$   $^1\text{H}$  NOESY : Proton proton nuclear Overhauser effect spectroscopy  
 $^1\text{H}$   $^1\text{H}$  TOCSY : Proton proton total correlation spectroscopy  
 $^1\text{H}$   $^{15}\text{N}$  HSQC : Proton nitrogen heteronuclear single quantum correlation  
 $^1\text{H}$   $^{15}\text{N}$  ROESY : Proton nitrogen rotational nuclear Overhauser effect spectroscopy  
4-HCCA : 4-hydroxy- $\alpha$ -sinamic acid  
ABTS : 2,2'-azino-bis(3)-ethylbenzthiazoline-6-sulphonic acid  
ADN : Acide désoxyribonucléique  
ARN : Acide ribonucléique  
BMP receptor : Bone morphogenic protein receptor  
BPTI : Bovin pancreatic trypsin inhibitor  
BSA : Bovin serum albumin  
CAM : Carboxylamideméthyl  
CNBr : Cyanogen bromide  
CIP : Calf intestine phosphatase  
DAB : Diaminobenzidine  
DMSO : Dimethyl sulfoxide  
DTT : Dithiothréitol  
Eau PPI : eau pour préparations injectables  
EDTA : acide éthylène-diamine-tétraacétique  
EGF : Epidermal growth factor  
ELISA : Enzyme-linked immunosorbent assay  
ESRF : European synchrotron radiation facility  
FPLC : Fast protein liquid chromatography  
GAR-PO : Goat Anti-Rabbit PerOxydase  
GIRK : G-protein activated inwardly rectifying K<sup>+</sup> current  
GSH, GSSG : Glutathion réduit, glutathion oxydé  
HCl : Chlorhydric acid  
HPLC : High performance liquid chromatography  
HSP : Heat shock protein  
ICl : Iodine monochloride  
IgG : Immunoglobulines de classe G  
IPTG : isopropyl- $\beta$ -D-thiogalactopyranoside  
KLH : Keyhole limpet hemocyanin  
LCI : Leech carboxypeptidase inhibitor  
lk1, lk2, lk3 : linker 1, linker 2, linker3  
MALDI TOF : Matrix assisted laser desorption ionisation - Time of flight  
MAP kinase : Mitogen activated protein kinase  
Milieu LB : milieu Luria Broth  
MTSBS : Méthane-thio sulfonato butyl sulfonate  
MPD : 2,4-methyl-pentane-diol  
NaCl : Chlorure de sodium  
NCBI : National center for biotechnology information  
nOe : Nuclear Overhauser effect



PBS : Phosphate buffered salin  
PCR : Polymerase chain reaction  
PDB : Protein database  
PEG 6000 : Polyethylen glycol 6000 Da  
PEI : Polyethyleneimine  
PMSF : Phenylmethylsulphonyl fluoride  
PVDF : Polyfluorure de vinylidène  
RIA : Radio immuno assay  
RMN : Résonance magnétique nucléaire  
RMSD : Root mean square distance  
RNase A : Ribonuclease A  
RU : Resonance unit  
SDS-PAGE : Sodium-dodécyl-sulfate – Polyacrylamide gel electrophoresis  
TAP : Tick anticoagulant protein  
TCEP : tris(2-carboxyethyl)phosphine  
TFA : Trifluoroacetic acid  
TGF beta receptor : Transforming growth factor beta receptor  
Tox60, tox61, tox62 : Toxine  $\alpha$ 60, toxine  $\alpha$ 61, toxine  $\alpha$ 62  
Tox61pT3, Tox62pT3 : Toxine  $\alpha$ 61 pont 3, toxine  $\alpha$ 62 pont 3  
TSP : 3-(trimethylsilyl)[2,2,3,3-2H<sub>4</sub>] propionate  
UV: Ultra violet

# INTRODUCTION



---

# 1 LE REPLIEMENT DES PROTEINES : UN DEFI MAJEUR DE LA RECHERCHE EN BIOLOGIE

---

## 1.1 De l'intérêt d'étudier le repliement des protéines

Les protéines furent découvertes par le chimiste hollandais Gerardus Johannes Mulder (1802-1880). Le terme « protéine » vient du grec ancien *prōtos* qui signifie premier, essentiel. Les protéines sont en effet des constituants essentiels des êtres vivants et représentent généralement environ la moitié du poids sec des cellules. Une autre étymologie, selon laquelle « protéine » ferait référence au dieu grec *Protée* qui avait le pouvoir de changer de forme à volonté, évoque une autre propriété fondamentale des protéines : ces molécules adoptent de multiples formes (ou structures) qui leur permettent d'assurer de multiples fonctions.

Une protéine est un polymère linéaire issu de la combinaison de 20 acides aminés. La séquence de ces acides aminés est déterminée par un gène porté par une séquence d'ADN. La production d'une protéine à partir d'un gène résulte de l'application du code génétique ((Jacob and Monod, 1961), (Crick et al., 1961), (Matthaei and Nirenberg, 1961)). Les processus biologiques assurant l'application de ce code sont la transcription de l'ADN en Acide RiboNucléique (ARN) et la traduction de ce dernier en protéine. Lors de la traduction, les protéines sont synthétisées par des ribosomes à partir d'un ARN messenger. La chaîne polypeptidique synthétisée comme une molécule linéaire acquiert une structure tridimensionnelle lui conférant sa capacité à réaliser une fonction : ce phénomène est appelé repliement des protéines. Les expériences menées par Anfinsen à la fin des années 50 et celles de Merrifield à la fin des années 60 ((Haber and Anfinsen, 1961), (White, 1961), (Guthe and Merrifield, 1968)) ont démontré que la structure native est entièrement déterminée par la nature et l'ordre (autrement dit la séquence) des acides aminés constituant la protéine. Cependant le mécanisme par lequel les protéines acquièrent leur structure demeure une énigme. La question de -comment une chaîne polypeptidique atteint sa structure tridimensionnelle ? – est un défi majeur de la recherche en biologie.

Outre l'intérêt fondamental de cette question, l'apport d'une réponse rendrait possible la prédiction de la structure d'une protéine à partir de sa séquence en acides aminés. Aujourd'hui, les génomes de 4450 espèces ont été séquencés (banque entrez Genome du NCBI : <http://www.ncbi.nlm.nih.gov>), chaque séquençage produit des dizaines de milliers de gènes correspondant à autant de séquences protéiques. La résolution de la structure de toutes ces protéines par des méthodes biophysiques traditionnelles étant impossible à cette échelle, une prédiction fiable *in silico* serait utile à l'interprétation biologique de ces données génétiques. Ce type de prédiction doit s'appuyer sur une théorie des processus de repliement. D'autre part une meilleure connaissance de ces processus présenterait un intérêt technologique permettant d'améliorer le rendement des méthodes de production des protéines par voie recombinante. En effet ces rendements sont souvent limités par la formation de molécules mal repliées qui ont tendance à précipiter((Graslund et al., 2008)). La compréhension des mécanismes de repliement constituerait dans ce cas une base à l'ingénierie du mécanisme de repliement des protéines recombinantes.

## 1.2 L'expérience originelle

Dans une série d'expériences décrites dans des articles parus à la fin des années 50 et au tout début des années 60, Anfinsen et ses collaborateurs ont été les premiers à étudier le repliement des protéines. Ils ont utilisé la Ribonucléase A, une petite protéine riche en ponts disulfure, comme protéine modèle. Dans leur expérience, les auteurs ont dénaturé et réduit cette protéine puis ont remplacé la solution de dénaturation par une solution renaturante. Leurs résultats montrent que spontanément la protéine retrouve son activité enzymatique ((Haber and Anfinsen, 1961)) et sa structure secondaire et tertiaire ((White, 1961)). La démonstration était faite que l'information nécessaire à l'acquisition de la structure tridimensionnelle des protéines est contenue dans leur séquence en acides aminés.

## 1.3 Théories décrivant le repliement des protéines

Depuis les travaux d'Anfinsen *et al.*, de nombreuses théories ont visé à proposer une explication générale aux phénomènes observés lors du repliement des protéines.

### 1.3.1 Paradoxe de Levinthal

Après l'observation de Anfinsen *et al.* selon laquelle les protéines se replient par elles-mêmes, suivait naturellement la question : comment les protéines « trouvent » leur forme native ? Au début des années 60, Cyrus Levinthal a proposé un raisonnement faisant la démonstration de la difficulté de cette question ((Levinthal, 1969)). Pour schématiser son raisonnement considérons un polypeptide constitué de 100 acides aminés. En supposant que chaque acide aminé peut adopter deux conformations (ce qui est largement inférieur à la réalité), le nombre total de configurations du polypeptide est alors égal à  $2^{100}$  soit  $10^{30}$ . Si l'on fait l'hypothèse que le temps de conversion d'une configuration à l'autre est de l'ordre de  $10^{-13}$  secondes, le temps nécessaire à l'exploration de toutes les configurations est de  $10^{30} \times 10^{-13}$  soit  $10^{17}$  secondes ou environ un milliard d'années. Ce raisonnement suggère qu'il est impossible que les protéines se replient en explorant au hasard l'ensemble de l'espace conformationnel possible. Les protéines doivent emprunter des voies préférentielles qui les guident vers la structure native.

### 1.3.2 Modèle de nucléation

A la suite de ces réflexions, les scientifiques s'intéressant à la question du repliement des protéines ont cherché un mécanisme simple capable d'expliquer ce paradoxe. Le premier modèle qui a été proposé a été baptisé « modèle de nucléation » ((Wetlaufer, 1973)). Dans ce modèle un noyau de structure secondaire apparaît en un point de la séquence et le reste de la structure s'organise autour de ce noyau. Ce modèle prédit l'absence d'intermédiaires lors du repliement. A l'époque où il a été proposé, les études de repliement s'intéressaient principalement à l'apparition de ces intermédiaires ce qui explique la faible adhésion à ce modèle. Les modèles suivants, proposés peu après, ont prévalu.

### 1.3.3 Framework model

La nucléation intervient en plusieurs points de la chaîne polypeptidique et génère des structures secondaires qui diffusent et se rassemblent pour former des microstructures présentant une conformation native. La vitesse d'apparition et de disparition de ces structures serait contrôlée par la vitesse de diffusion dans le solvant ((Ptitsyn, 1987), (Ptitsyn and Rashin, 1975), (Kim and Baldwin, 1990)). Les arguments expérimentaux corroborant ce modèle provenaient de l'étude de peptides extraits de la séquence de différentes protéines qui avaient la capacité de former des structures secondaires par eux-mêmes (principalement des hélices  $\alpha$  et des coudes  $\beta$ ,

mais aussi parfois des  $\beta$ -hairpin) ((Brown and Klee, 1971), (Bierzynski et al., 1982), (Shoemaker et al., 1985)). Ces structures constitueraient des points d'initiation du repliement. Cependant, il a été par la suite observé que la plupart des peptides qui forment des structures secondaires dans les protéines entières sont déstructurées une fois extraits du reste de la séquence ((Eppand and Scheraga, 1968)). Il semble donc que les interactions hydrophobes soient indispensables à la stabilisation des structures secondaires. Le modèle de l'« hydrophobic collapse » traduit en français par effondrement hydrophobe a été proposé.

### **1.3.4 Effondrement hydrophobe**

Le premier événement serait une modification de la solvation de la molécule induisant son effondrement désordonné. Le rayon hydrodynamique diminuerait rapidement pour donner un globule déstructuré au sein duquel des réarrangements permettraient l'apparition de structures secondaires puis de la structure native ((Schellman, 1955), (Kauzmann, 1959), (Tanford, 1962)). Ce modèle a été proposé suite à la découverte selon laquelle la force motrice principale du repliement des protéines est fournie par l'expulsion de l'eau liée aux chaînes latérales hydrophobes lorsque ces dernières s'associent et s'isolent du solvant ((Kauzmann, 1959)). Cependant ce modèle pose un problème théorique car il paraît peu probable que la formation de nombreuses interactions non spécifiques soit favorable à un réarrangement ultérieur de la protéine.

### **1.3.5 Le globule fondu (molten globule en anglais)**

La théorie du globule fondu permet de réconcilier les deux propositions ((Ptitsyn and Rashin, 1975), (Ptitsyn, 1995)). Dans ce modèle, l'« hydrophobic collapse » et la formation des premières structures secondaires ont lieu simultanément pour former le globule fondu qui lui, a la capacité de se réarranger pour générer la forme native. Uversky et Fink ont analysé les résultats obtenus lors de l'étude du repliement de 41 protéines et ont mis en évidence que la diminution du volume hydrodynamique et l'apparition de structures secondaires sont simultanées ((Uversky and Fink, 2002)).

### **1.3.6 L'entonnoir énergétique de repliement (Funnel theory)**

Plus récemment, grâce au développement de l'informatique et à l'augmentation des capacités de calcul, il est devenu envisageable de modéliser le repliement des protéines à partir des forces d'interaction entre atomes décrites par les lois de la Physique ((Fersht and Daggett, 2002)). En étudiant les caractéristiques énergétiques du repliement des protéines, les modélisateurs ont observé que le paysage énergétique des protéines naturelles a une forme d'entonnoir ((Bryngelson and Wolynes, 1987), (Leopold et al., 1992)).

L'étude des protéines naturelles montre que ces polymères ont la capacité, à partir d'une séquence, de former systématiquement la même structure. Le fait que cette structure soit spécifique de la molécule provient de l'hétérogénéité de la chaîne polypeptidique. La proximité ou l'éloignement de deux résidus implique l'existence d'états énergétiques différents, les deux conformations associées sont alors plus ou moins stables. Dans le cas d'une protéine artificielle dont la séquence est choisie au hasard, les minimums énergétiques peuvent correspondre à différentes structures ((Hardin et al., 2002)). L'état d'énergie minimum provient alors du résultat aléatoire de la compétition qui existe entre les différentes contributions énergétiques. Un changement minime dans la séquence de cette protéine (comme une mutation ponctuelle) provoquerait une modification importante de la structure du minimum d'énergie. Ce type de comportement serait très préjudiciable à un organisme vivant puisque si la molécule ancestrale assurait une certaine fonction, la perte de la structure suite à la mutation empêcherait la molécule descendante d'assurer cette fonction. Les processus de sélection engendrent donc des séquences dans lesquelles les interactions ne sont pas en conflit avec la structure préexistante mais qui au

contraire permettent d'atteindre de manière coopérative un état de frustration minimal (« minimal frustration » en anglais, un concept qui provient de la théorie des verres de spin en physique (Bryngelson and Wolynes, 1987)). Au début du repliement, la population de protéines dépliées est hétérogène et occupe un vaste ensemble de configurations de haute énergie (la partie évasée de l'entonnoir). Lors du repliement, les molécules descendent la pente énergétique, évoluent vers des états de moindre énergie, l'espace conformationnel possible est de plus en plus restreint et finalement converge vers l'état natif (la partie étroite de l'entonnoir). (Figure 1 extraite de (Dobson et al., 1998)).

Ce concept a été récemment confirmé par l'expérience par Baker *et al.*. En effet ces auteurs sont parvenus à concevoir par simulation numérique puis à produire expérimentalement un nouveau domaine protéique qu'ils ont baptisé Top7 ((Kuhlman et al., 2003)). Cette protéine artificielle est plus stable que la plupart des protéines de même taille. Ces mêmes auteurs ont montré qu'il existe au moins trois étapes dans son repliement, qu'une conformation non native est stable à l'équilibre et que de nombreux fragments isolés de la séquence de Top7 sont stables ((Watters et al., 2007)). Ces résultats indiquent que le repliement de Top7 est beaucoup moins coopératif que celui de protéines naturelles de même taille. Ce résultat tend à prouver que l'aspect coopératif du repliement (et donc l'existence d'un entonnoir énergétique lisse) n'est pas une propriété générale à toutes les chaînes polypeptidiques mais au contraire le produit de la sélection naturelle.

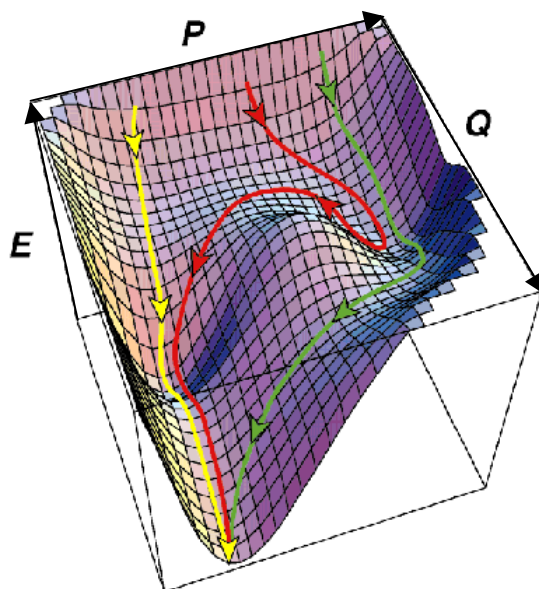


Figure 1: Schéma représentant la théorie de l'entonnoir. E : énergie, Q : proportion de contacts natifs, P : mesure de l'espace conformationnel disponible ((Dobson et al., 1998)).

La théorie de l'entonnoir modélise l'aspect coopératif du repliement souligné par Levinthal tout en proposant une explication évolutive au fait que ce comportement est partagé par la majorité des protéines naturelles étudiées. Toutes les théories proposées auparavant sont compatibles avec ce modèle, puisque dans le contexte de la théorie de l'entonnoir, il n'existe pas de différence fondamentale entre les interactions impliquées dans les structures secondaires et celles participant à structure tertiaire. Ce modèle prédit que plusieurs voies de repliement peuvent coexister dans le paysage énergétique d'une même protéine (Figure 1). Dans le cadre de cette théorie, deux séquences très similaires voire identiques, peuvent suivre des chemins de repliement tout à fait différents.

## 1.4 Méthodes expérimentales de caractérisation du chemin de repliement des protéines sans ponts disulfure

Historiquement, les premières études du repliement des protéines se sont focalisées sur les protéines à ponts disulfure (voir partie 3 de cette introduction pour le détail de la méthodologie associée à ce type d'étude). Plus récemment le développement de deux méthodes a permis d'obtenir des informations sur le repliement des protéines sans ponts disulfure.

### 1.4.1 Caractérisation de l'état de transition par la méthode des « $\phi$ -value »

Lorsqu'une protéine sans pont disulfure est placée dans un tampon contenant une concentration suffisamment importante de dénaturant chimique, la protéine se dénature. Ce processus (comme le processus inverse de renaturation lorsque la protéine dénaturée est placée dans un tampon sans dénaturant chimique) s'avère rapide dans la majorité des cas (entre la milliseconde et la seconde : (Basharov, 2003)) et procède souvent selon un mécanisme à deux états. Ce processus peut cependant être suivi grâce à l'utilisation de systèmes expérimentaux de type « stopped flow » ((Roder et al., 2004)). Ce type d'expérience permet de déterminer l'énergie d'activation associée à la réaction de repliement. Fersht *et al.* ont proposé dans une série d'articles parue en 1992 une stratégie permettant de caractériser structuralement l'état de transition associé à la réaction de repliement ((Fersht et al., 1992), (Serrano et al., 1992a), (Serrano et al., 1992b), (Matouschek et al., 1992a), (Matouschek et al., 1992b), (Serrano et al., 1992c)). Pour appliquer cette méthode, il est nécessaire de produire des mutants ponctuels de la protéine étudiée puis de déterminer les énergies d'activation associées à leurs réactions de repliement. Pour chacun des mutants, un critère numérique appelé «  $\phi$ -value » est calculé pour comparer l'énergie d'activation du mutant à celle de la protéine native. L'interprétation des valeurs de «  $\phi$ -value » obtenues pour chaque mutant permet de savoir si le résidu muté est structuré dans l'état de transition comme dans la molécule native ou non. L'analyse pour une protéine, d'un ensemble de «  $\phi$ -value » donne une image structurale de l'état de transition.

### 1.4.2 Méthode d'analyse du repliement par des expériences d'échange de deutérium

Le suivi de l'échange des protons amides des protéines en utilisant le deutérium comme marqueur et en analysant cet échange par RMN (Résonance Magnétique Nucléaire) a permis de détecter et caractériser des intermédiaires apparaissant lors du repliement des protéines sans ponts disulfure ((Krishna et al., 2004)). En réalisant des expériences de pulse d'échange d'hydrogène, il est possible de caractériser la structure, la stabilité et la vitesse de repliement d'intermédiaires qui existent pendant des temps inférieurs à la seconde. Dans ce type d'expérience, la protéine est dans un premier temps dénaturée dans une solution de deutérium. Tous les protons amide de la molécule sont alors remplacés par du deutérium. Le repliement est déclenché par une dilution rapide (« stopped flow ») dans un tampon permettant le repliement préparé avec de l'eau et dont le pH est faible afin d'éviter l'échange pendant cette phase. Après un certain temps de repliement (noté Tf) dans ces conditions un pulse de marquage à l'hydrogène est effectué en augmentant brusquement le pH. Les fonctions amides encore accessibles au solvant s'échangent mais celles déjà enfouies dans des structures sont protégées. Dans une dernière étape le pH est à nouveau abaissé dans ce même milieu de repliement afin de figer le marquage effectué au sein de la structure finale de la molécule. Des expériences de RMN permettent ensuite de déterminer les protons protégés et ceux qui ne le sont pas après un temps Tf de repliement.



## 2 LE REPLIEMENT COTRADUCTIONNEL DES PROTEINES

Comme on l'a vu précédemment, au sein des cellules des organismes vivants, les protéines sont synthétisées par les ribosomes. Une fois qu'une partie des résidus émerge du ribosome, la chaîne polypeptidique naissante peut commencer à se replier. Les protéines étant synthétisées depuis l'extrémité N-terminale vers l'extrémité C-terminale, il est possible que le repliement soit initié dans la partie N-terminale alors même que le reste de la molécule n'a pas encore été synthétisé. Cette situation est différente de celle observée dans les expériences de repliement réalisées *in vitro* par renaturation d'une protéine dénaturée car dans ce dernier cas l'initiation peut avoir lieu dans n'importe quelle partie de la séquence. Les mécanismes de repliement *in vivo* et *in vitro* sont potentiellement très différents du fait de la vectorisation du repliement imposé par l'aspect cotraductionnel du repliement.

### 2.1 Arguments contre

Plusieurs observations tendent à prouver que le repliement des protéines est un phénomène indépendant de la synthèse peptidique :

- De nombreuses protéines ont la capacité, après dénaturation, d'acquies leur structure par elles-même ((Haber and Anfinsen, 1961), (White, 1961)). La structure est uniquement déterminée par la séquence en acides aminés.
- Il a été montré pour quelques protéines que leur structure finale n'était pas sensible à la permutation de la séquence ((Alexandrov, 1993)). Dans ces expériences les extrémités de ces protéines étaient mises bout à bout artificiellement puis coupées à différentes positions dans la séquence circulaire.
- Certaines protéines n'ont pas la capacité de se replier lorsque leur partie C-terminale est supprimée ((Tsou, 1988)).

### 2.2 Arguments pour

#### 2.2.1 Preuves indirectes

##### 2.2.1.1 Comparaison vitesse de repliement, vitesse de traduction

Lors du repliement d'une protéine, la condensation de la molécule et les structures secondaires apparaissent en moins d'une seconde ((Roder and Colon, 1997)). Sachant que la vitesse de synthèse peptidique est de l'ordre de 50 à 300 résidus par minute *in vitro* et est légèrement plus rapide *in vivo* ((Fedorov and Baldwin, 1998)), les premières étapes du repliement peuvent avoir lieu pendant la synthèse des protéines((Fedorov and Baldwin, 1998), (Basharov, 2003)). Cette observation constitue une preuve indirecte du repliement cotraductionnel des protéines.

##### 2.2.1.2 Biais structural lié au repliement cotraductionnel

A partir de l'analyse des structures d'une banque de données contenant 215 structures de protéines, Alexandrov a montré qu'en moyenne, les résidus formaient plus de contacts avec les résidus situés en amont dans la séquence (du côté N-terminal) qu'avec ceux situés en aval ((Alexandrov, 1993)). Il a proposé que cette observation puisse constituer un argument en faveur de l'existence généralisée d'un repliement cotraductionnel. Il déduit de cette observation que la partie N-terminale des structures protéiques devait nécessairement être plus compacte que la partie C-terminale sans pour autant en apporter la preuve.

Plus récemment Laio et Micheletti ont obtenu la conclusion inverse (à savoir qu'en général, la partie C-terminale est plus compacte que la partie N-terminale) à partir de l'analyse de 2000 protéines non-homologues((Laio and Micheletti, 2006)).

Dernièrement Deane *et al.* ont cherché à mettre en évidence l'aspect cotraductionnel du repliement à partir de l'analyse statistique de 4300 structures de protéines ((Deane et al., 2007)). D'après ces auteurs, si les protéines se replient au cours de leur synthèse sur le ribosome ce phénomène devrait influencer la structure finale de la molécule. D'une part, la partie N-terminale devrait être plus enfouie que la partie C-terminale, d'autre part, on s'attend à ce que les acides aminés produisent plus de contact avec les résidus situés en amont dans la séquence qu'avec ceux situés en aval. Ils ont mis au point trois critères pour tester:

- le rapport des distances moyennes des segments N et C-terminaux par rapport au centre de gravité de la structure,
- la distance entre le segment N-terminal et le résidu le plus proche du centre de gravité,
- une mesure équivalente à celle proposée par Alexandrov, pour déterminer si les acides aminés établissent en moyenne plus de contacts avec les résidus situés en amont dans la séquence.

Leurs résultats montrent que pour les structures des protéines appartenant à la famille  $\alpha/\beta$  de la banque SCOP, les trois critères vont dans le sens d'un effet de l'aspect cotraductionnel du repliement sur la structure finale. Pour les autres classes, la conclusion est moins nette mais des signes de repliement cotraductionnel sont observables.

## 2.2.2 Preuves expérimentales directes

### 2.2.2.1 Mise en évidence de la formation cotraductionnelle de ponts disulfure :

Trois articles décrivent la formation d'un pont disulfure de manière cotraductionnelle dans trois protéines modèles différentes :

Bergman et Kuehl, en 1979 ((Bergman and Kuehl, 1979)), ont été les premiers à mettre en évidence la formation d'un pont disulfure de manière cotraductionnelle. Leur modèle d'étude était la chaîne légère de l'immunoglobuline de souris qui comprend un pont disulfure intramoléculaire et une cystéine libre permettant de former un pont avec la chaîne lourde. La procédure suivie était la suivante : ils disposaient de cellules de souris produisant en abondance la chaîne légère d'immunoglobuline, ces cellules étaient lysées, puis une fraction contenant uniquement les chaînes peptidiques en cours de synthèse était obtenue selon un procédé développé auparavant. Toutes ces étapes étaient réalisées en présence de iodoacétamide afin de bloquer les cystéines libres et d'éviter la formation de ponts disulfure pendant le traitement. La fraction obtenue était ensuite réduite puis les cystéines libérées par la réduction étaient bloquées à l'acide iodoacétique marqué au tritium. Ensuite une étape d'immunoprécipitation permettait de récupérer uniquement les chaînes légères d'immunoglobuline. La solution de chaînes naissantes d'immunoglobuline était ensuite séparée par filtration sur gel. Les résultats montrent une incorporation d'acide iodoacétique tritié et donc la présence d'un pont disulfure dans les chaînes naissantes. Une digestion trypsique de ces dernières a permis de mettre en évidence que les cystéines impliquées dans ce pont formé de manière cotraductionnelle correspondent aux cystéines du pont natif.

Plus récemment, Chen *et al* ont mis en évidence l'apparition d'un pont disulfure de manière cotraductionnelle dans l'hémagglutinine A du virus de la grippe ((Chen et al., 1995)). Dans cette étude, les auteurs ont utilisé des cellules d'ovaires de hamster chinois (CHO) infectées par le virus de la grippe. Un marquage instantané à la cystéine et à la méthionine  $^{35}\text{S}$  de 1 minute 30 était effectué. Ce temps de marquage est inférieur à celui de la synthèse complète de l'hémagglutinine A qui dure 2 minutes dans ce système. Les cellules étaient ensuite traitées au N-éthylmaléimide (un agent alkylant qui traverse les membranes) afin d'éviter toute oxydation pendant le traitement. Les chaînes naissantes et les molécules d'hémagglutinine complètement synthétisées étaient séparées du reste des extraits cellulaires par immunoprécipitation grâce à des anticorps dirigés contre la partie N-terminale de la molécule. Le mélange de chaînes naissantes obtenu était ensuite analysé par SDS-PAGE bidimensionnelle. La première dimension était réalisée en condition non

réductrice et la seconde en condition réductrice. La Figure 2 présente le principe de cette méthode :

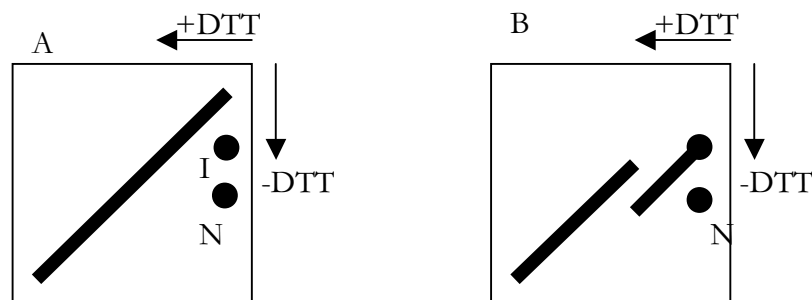


Figure 2: SDS-PAGE bidimensionnelle pour l'étude de la formation cotraductionnelle de ponts disulfure. A : formation post-traductionnelle des ponts, B : formation cotraductionnelle d'un pont

En absence de ponts disulfure les polypeptides vont migrer selon leur taille dans les deux directions. La protéine en cours de synthèse se distribuera donc sur une diagonale où le point inférieur gauche marque l'emplacement des peptides les plus courts, et le coin supérieur droit la protéine entière (Figure 2-A). L'apparition d'un pont disulfure va augmenter le degré de compacité du polypeptide considéré et induire une migration plus rapide en absence de réducteur entraînant la formation d'une discontinuité dans la diagonale (Figure 2-B). La localisation de cette discontinuité permet une identification approximative du pont formé.

#### 2.2.2.2 Mise en évidence de l'apparition cotraductionnelle d'un épitope :

Il a été montré par la liaison d'un anticorps spécifique de la forme repliée que la chaîne  $\beta$  de la tryptophane synthase d'*Escherichia coli* commence à se replier lors de sa synthèse ((Fedorov et al., 1992), (Friguet et al., 1993), (Tokatlidis et al., 1995)). La même équipe a mis en évidence l'apparition cotraductionnelle d'un épitope dans la protéine « tailspike » du phage P22 grâce à un anticorps monoclonal ( (Friguet et al., 1994)).

#### 2.2.2.3 Mise en évidence de la formation cotraductionnelle d'un domaine catalytique :

L'étude menée par Nicola *et al.* montre que le repliement cotraductionnel est possible lors de la synthèse protéique dans le cytosol de cellules CHO ((Nicola et al., 1999)). Ces auteurs ont choisi comme modèle une protéine de la capsid d'un alpha virus : le virus de la forêt Semliki (Semliki Forest virus capsid C protein). Les cinq protéines de la capsid de ce virus sont produites simultanément en un seul précurseur à partir d'un même ARNm. La protéine C est située en partie N-terminale de ce précurseur. Les auteurs ont montré que cette protéine a une activité autocatalytique et que la séparation de la protéine C a lieu avant que le reste du précurseur ne soit synthétisé.

#### 2.2.2.4 Mise en évidence de l'augmentation de l'efficacité de repliement des protéines multidomaines :

Netzer et Hartl ((Netzer and Hartl, 1997)) sont partis de l'observation que *in vitro* le repliement des protéines multidomaine est souvent peu efficace et conduit à leur agrégation. Ils ont utilisé une protéine artificielle obtenue par la fusion de deux protéines mono-domaine reliées par une séquence peptidique correspondant à celle d'un « linker » de protéines naturelles. Le repliement *in vitro* de cette protéine synthétique testé par dénaturation chimique puis dilution dans un tampon de repliement aboutissait à un rendement quasi-nul. Parallèlement les auteurs ont étudié l'expression de cette protéine dans un système eucaryote de traduction *in vitro*. Leurs résultats montrent qu'avec ce système le rendement de repliement est proche de 100% et que le domaine

N-terminal se forme de manière cotraductionnelle. Leurs résultats indiquent que non seulement le repliement cotraductionnel existe mais qu'il peut s'avérer nécessaire à un repliement efficace dans le cas de certaines protéines multidomaines.

La luciférase est une protéine luminescente composée de deux domaines qui a été utilisée comme modèle pour étudier le repliement cotraductionnel par Frydman *et al.* ((Frydman et al., 1999)). Ces auteurs ont produit cette protéine par synthèse ribosomale *in vitro* et ont observé l'accumulation d'un intermédiaire de repliement dans lequel la partie N-terminale de la molécule est repliée. A l'inverse, lorsqu'ils ont dénaturé puis replié cette protéine *in vitro*, le rendement de repliement était beaucoup plus faible et l'intermédiaire structuré dans la partie N-terminal n'apparaissait pas. Cependant ces résultats ne permettaient pas d'être certain que le gain d'efficacité observé lors de la synthèse ribosomale *in vitro* était uniquement dû à l'aspect cotraductionnel du repliement et non à la présence de protéines chaperon. Les expériences menées par Svetlov *et al.* ((Svetlov et al., 2006)) ont levé cette ambiguïté concernant les protéines chaperons de la famille HSP70. Ces auteurs ont étudié le repliement de la luciférase à partir de son état dénaturé (repliement *in vitro*) et ont montré que l'adjonction de protéines chaperons de type HSP70 permet de passer d'un rendement 8 à 80%. Ils ont alors utilisé un système de traduction *in vitro* pour produire la luciférase et dans ce cas l'adjonction des chaperons de la famille HSP70 ne modifie pas le rendement. Leur expérience montre que le gain de rendement obtenu lors du repliement cotraductionnel ne nécessite pas l'action de protéines chaperons de la famille HSP70.

## 2.3 Simulation du repliement cotraductionnel

La plupart des études de modélisation du repliement des protéines ont implicitement supposé que le point de départ du repliement est une molécule dénaturée et présente dans son intégralité. Pourtant les travaux présentés ci-dessus ont suggéré la possibilité que les protéines commencent à se replier alors qu'elles n'ont pas encore été entièrement synthétisées. Récemment, plusieurs équipes ont tenté de tenir compte dans leur modélisation de l'aspect cotraductionnel du repliement.

### 2.3.1 Par génération séquentielle

Morrissey *et al.* ((Morrissey et al., 2004)) ont réalisé une simulation de repliement de type « lattice model » à trois dimensions en utilisant un potentiel de Gō. Les mouvements des « molécules » (il s'agit à proprement parler de modèles de squelette peptidique) étaient simulés grâce à un algorithme de type Monte Carlo. Les auteurs ont modélisé l'aspect vectoriel du repliement en partant d'un résidu et en simulant l'addition récursive d'un « résidu » après un certain temps de mise à l'équilibre. Ils ont comparé les résultats obtenus à ceux d'une simulation du repliement à partir de l'état dénaturé. Deux types de molécules ont été testés : la première contenait uniquement des contacts locaux et simulait donc le comportement d'une protéine riche en hélice  $\alpha$ , la deuxième était choisie afin de minimiser les contacts locaux et s'apparentait donc plutôt à une protéine riche en feuillets  $\beta$ . Leurs résultats montrent que la simulation cotraductionnelle aboutit à une cinétique de repliement plus rapide que la simulation contrôle et ce, plus nettement dans le cas de la molécule  $\alpha$  contenant de nombreux contacts locaux. Leurs résultats indiquent également qu'en passant d'un repliement cotraductionnel au repliement contrôle il est possible d'obtenir deux états natifs différents à partir de la même séquence de départ. Huard *et al.* ont mené le même type d'étude en utilisant un « lattice model » HP (c'est à dire en considérant que la séquence contient deux types de monomères, soit hydrophobe, soit hydrophile) à deux dimensions ((Huard et al., 2006)). L'élongation de la chaîne était modélisée de la même façon que dans l'étude de Morrissey *et al.*. La simulation des mouvements de la chaîne

était réalisée par un modèle déterministe. Par cette méthode les auteurs ont montré que dans certaines conditions de contrôle cinétique, la structure finale obtenue peut ne pas être la structure de plus basse énergie. Les résultats de ces deux études indiquent que la prise en compte de l'aspect cotraductionnel du repliement lors de simulations pour la prédiction de structure peut avoir une influence sur la structure finale obtenue.

Wang *et al.* ont également proposé une simulation du repliement cotraductionnel avec un « lattice model » en utilisant deux séquences protéiques de 27 et 36 résidus ; les mouvements étant modélisés par Monte Carlo ((Wang and Klimov, 2008)). Ils ont comparé leurs résultats à la simulation du repliement de la même molécule effectuée à partir d'un état dénaturé. Leur conclusion était que les trajectoires de repliement sont identiques que l'on tienne compte ou non de l'aspect vectoriel.

### 2.3.2 Par extrusion à travers un canal

Dans le but de se rapprocher au maximum des conditions de repliement *in vivo* lors de la synthèse ribosomale, deux équipes ont simulé le repliement d'une chaîne polypeptidique émergeant d'un canal.

Kirmizialtin *et al.* ont simulé l'extrusion d'une séquence polypeptidique connue pour adopter une structure en  $\beta$ -hairpin à travers des canaux de différents diamètres ((Kirmizialtin et al., 2004)). Leur simulation utilise un modèle « off lattice » avec une séquence composée de trois types de monomères : hydrophobe, hydrophile ou neutre. La dynamique a été simulée par l'équation de Langevin. Deux régimes de repliement ont été observés, pour les grands diamètres la protéine se replie quelle que soit sa position dans le canal, pour les petits diamètres le peptide reste déplié. Le diamètre critique correspondant à la transition entre les deux régimes est proche de celui du canal du ribosome. Sachant que le diamètre de ce canal est variable, ce résultat suggère que l'apparition cotraductionnelle de structures dans le canal du ribosome ne serait pas universelle mais dépendrait de la séquence synthétisée.

Une simulation de la synthèse peptidique couplée à son extrusion à travers le canal d'échappement du ribosome a été proposée par Elcock ((Elcock, 2006)). L'auteur a utilisé un modèle de Gō pour simuler le repliement cotraductionnel de trois protéines modèles : l'inhibiteur de chymotrypsine II, la barnase et la protéine du semliki forest virus. La dynamique a été simulée par un algorithme de dynamique Brownienne. Le rapport entre les vitesses simulées de synthèse peptidique et de repliement a été fixé à partir des valeurs obtenues expérimentalement *in vitro* pour la barnase et l'inhibiteur de la chymotrypsine. La synthèse des résidus ainsi que l'extrusion ont directement été simulés au sein d'une structure de ribosome. L'auteur conclue que pour les deux protéines mono-domaine étudiées (à savoir la barnase et l'inhibiteur de la chymotrypsine), les trajectoires de repliement observées sont les mêmes dans la simulation cotraductionnelle et dans la simulation à partir de l'état dénaturé. A l'inverse la protéine du « semliki forest virus » qui est quant à elle composée de deux domaines, suit une trajectoire de repliement différente lorsqu'on introduit la possibilité du repliement cotraductionnel. La simulation de son repliement à partir d'un état dénaturé montre que les deux domaines se replient indépendamment alors que lors de la simulation cotraductionnelle, le domaine C-terminal vient se replier contre un domaine N-terminal déjà préformé.

Comme on vient de le voir, ce n'est que très récemment que des tentatives de prise en compte de l'aspect cotraductionnel dans les simulations de repliement sont apparues dans la littérature. En fonction du type de simulation effectuée et du modèle de protéine étudié, les conclusions divergent. Dans certains cas ((Wang and Klimov, 2008)), l'aspect cotraductionnel du repliement n'a pas d'effet sur le mécanisme de repliement, alors que dans d'autres, la vitesse de repliement voir même l'état final prédit peuvent être modifiés ((Morrissey et al., 2004), (Huard et al., 2006)). On peut espérer qu'à l'avenir les modélisateurs entreprennent l'étude simultanée, avec la même méthode, d'un grand nombre de protéines différentes.

---

### 3 LE REPLIEMENT OXYDANT *IN VITRO* DES PROTEINES A PONTS DISULFURE

---

Le terme « repliement oxydant » (oxidative folding en anglais) englobe l'ensemble des processus par lesquels une protéine à ponts disulfure réduite et dénaturée acquière à la fois sa structure native et les ponts disulfure correctement appariés qui correspondent à cette structure. Le phénomène d'oxydation des cystéines libres en un ensemble de ponts disulfure natifs est un phénomène complexe qui implique, en plus des modifications conformationnelles inhérentes au repliement, différentes réactions covalentes comme l'oxydation (formation d'un pont disulfure, Figure 3 A et B), la réduction (rupture d'un pont disulfure, Figure 3 A et B) et l'isomérisation (réarrangement d'un ou plusieurs ponts disulfure Figure 3 C). Malgré la complexité de ce mécanisme les peptides riches en ponts disulfure ont, depuis les premières expériences sur le repliement *in vitro* avec les travaux de Anfinsen ((White, 1961)(Haber and Anfinsen, 1961)), fréquemment été choisis comme modèles pour caractériser le ou les chemins qu'empruntent les protéines pour se replier. En effet, lors du suivi du repliement oxydant des protéines à ponts disulfure, ces ponts servent de sonde pour obtenir des informations structurales sur les étapes intermédiaires du processus suite à l'apparition d'espèces stables partiellement oxydées (appelées également intermédiaires de repliement) pouvant aisément être isolées et caractérisées. La formation ou la réduction de ces ponts est un phénomène localisé qui peut être interprété en terme d'acquisition de structure.

#### 3.1 La réaction élémentaire de formation d'un pont disulfure

##### 3.1.1 La réaction d'échange thiol/disulfure

La Figure 3 présente la réaction d'échange thiol/disulfure commune à toutes les réactions qui ont lieu lors du repliement oxydant. Un anion thiolate attaque la liaison disulfure pour former un intermédiaire dont la charge négative est délocalisée sur les trois atomes de soufre. La formation ou la réduction d'un pont disulfure par le glutathion oxydé ou réduit implique deux réactions de ce type (Figure 3 B). Lorsqu'un anion thiolate d'un résidu cystéine attaque un pont disulfure existant dans la protéine, on observe une réaction d'échange intramoléculaire conduisant au remaniement de l'appariement des ponts (Figure 3 C). Le terme anglais de « reshuffling » sera employé par la suite pour désigner ce phénomène. Ces figures sont adaptées de la revue de Narayan *et al.* ((Narayan et al., 2000)).

Pour toutes les réactions présentées en Figure 3 il est nécessaire de former un ion thiolate pour que la réaction d'oxydoréduction ait lieu. De ce fait, la réaction de repliement oxydant ne peut pas avoir lieu à pH acide car dans ce cas la chaîne latérale des cystéines est sous forme thiol.

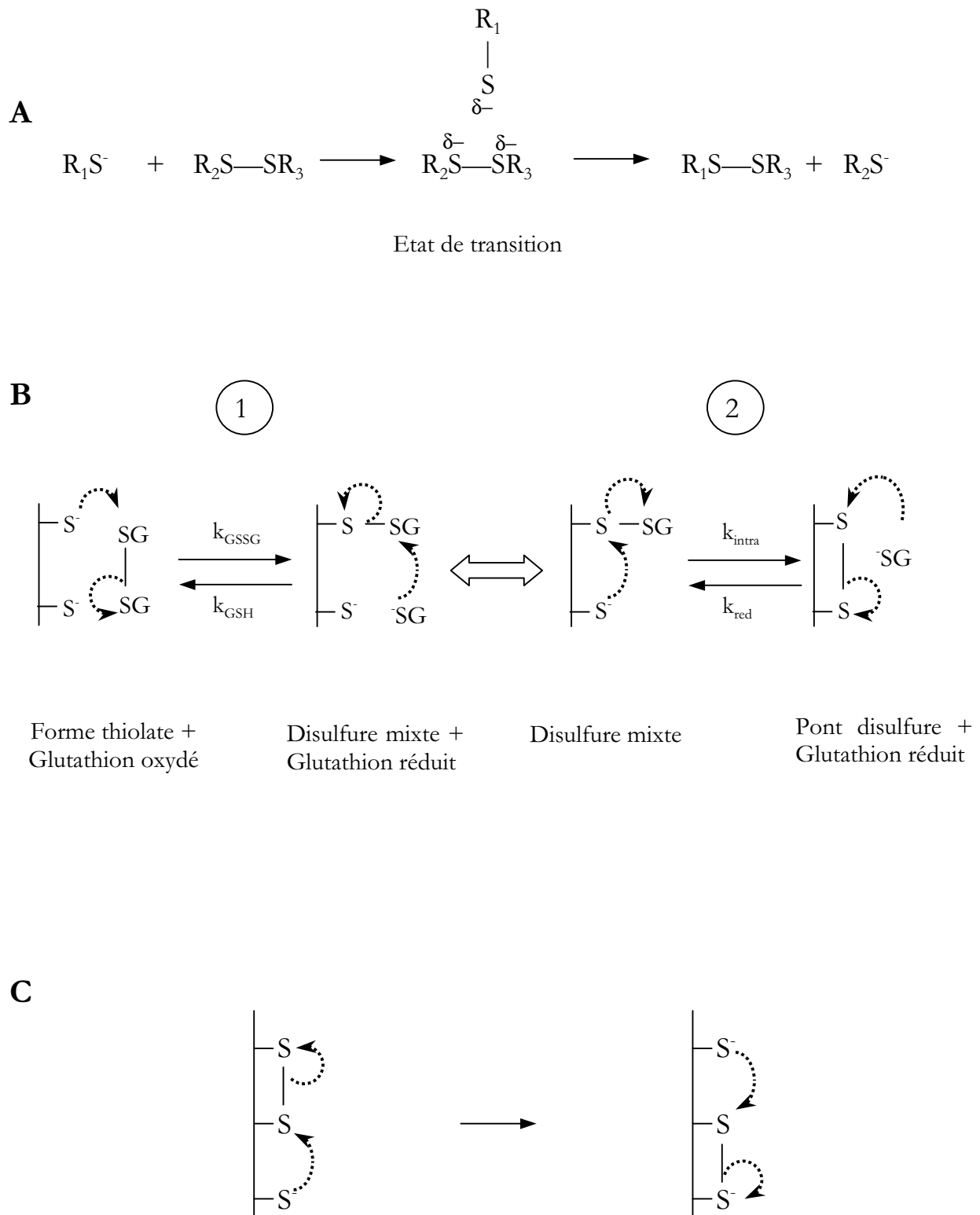


Figure 3 : Réactions chimiques impliquées dans le repliement oxydant des protéines à ponts disulfure.

### 3.1.2 Vitesse de formation d'un pont disulfure

Les informations présentées dans ce paragraphe ont été extraites de la revue suivante ((Gilbert, 1990)).

Comme cela est représenté sur la Figure 3 B, la réaction d'oxydation de deux cystéines pour former un pont disulfure se déroule en deux étapes. La première étape est l'attaque d'une molécule de glutathion oxydé par une cystéine libre dont le résultat est la formation d'un disulfure mixte. La constante de vitesse de cette réaction élémentaire est appelée  $k_{\text{GSSG}}$ . La deuxième étape consiste en l'attaque du disulfure mixte par la deuxième cystéine libre pour former le pont disulfure, la constante de vitesse correspondante est  $k_{\text{intra}}$ . La vitesse de formation du pont disulfure est imposée par l'étape la plus lente entre ces deux réactions, ou étape limitante.

La constante  $k_{\text{GSSG}}$  est fortement dépendante de l'accessibilité des cystéines libres. Lorsque ces dernières sont enfouies au sein de la structure de la protéine, dans une région inaccessible au solvant et donc a fortiori aux molécules de glutathion oxydé, la probabilité de rencontre du glutathion avec une cystéine libre est faible.

Pour que la deuxième étape de la réaction de formation d'un pont disulfure ait lieu, il faut qu'à proximité du disulfure mixte, la concentration locale en ions thiolates soit élevée. Deux facteurs principaux influent sur cette concentration :

- Les contraintes géométriques : Si, pour que la cystéine libre se trouve à proximité du disulfure mixte, la molécule doit passer par un état structural contraint (ou état « frustré » voir partie 1.3.6 de cette introduction), la concentration locale en ions thiolates à proximité du disulfure mixte sera faible.
- L'entropie conformationnelle :  $k_{\text{intra}}$  dépend de l'entropie de la forme présente avant la formation du pont. Si cette forme a la liberté d'explorer un large espace conformationnel, la probabilité de rencontre de la cystéine libre avec le disulfure mixte sera diminuée.



## 3.2 Les expériences de repliement oxydant

### 3.2.1 La cinétique de repliement

La première étape d'une expérience de repliement oxydant consiste à réduire et dénaturer (habituellement en présence de DTT réduit et de chlorure de guanidium) la protéine étudiée. La protéine ainsi dénaturée et réduite est séparée de son milieu de dénaturation puis placée dans un milieu permettant son repliement oxydant. Pour que la réaction de repliement ait lieu, il est nécessaire qu'un agent oxydant soit présent dans le milieu de repliement. Lors des premières expériences ((White, 1961), (Haber and Anfinsen, 1961)), le dioxygène contenu dans les solutions utilisées pour préparer le milieu de repliement suffisait pour que la réaction ait lieu. Par la suite, les expérimentateurs ont tenté de se rapprocher des conditions observées dans la cellule. Les protéines à ponts disulfure sont exportées dans le réticulum endoplasmique. Ce compartiment cellulaire contient le couple oxydoréducteur glutathion réduit / glutathion oxydé qui joue le rôle de tampon redox. Afin de se rapprocher de ces conditions, la plupart des études de repliement *in vitro* ont été menées en présence du couple glutathion réduit / glutathion oxydé (voir la formule développée du glutathion sur la Figure 4).

Dans l'article ((Rothwarf and Scheraga, 1993)), Scheraga et ses collaborateurs proposent quant à eux d'utiliser le couple DTT oxydé/DTT réduit (Figure 4) comme agent oxydo-réducteur. Selon eux cet agent oxydo-réducteur comporte deux avantages pour l'étude du repliement oxydant des protéines à ponts disulfure à savoir que :

- à la différence du glutathion, le DTT oxydé ne forme pas de disulfure mixte avec la protéine étudiée lorsqu'il est utilisé comme agent oxydant,
- toujours à la différence du glutathion, le potentiel redox du tampon redox DTT<sub>ox</sub>/DTT<sub>red</sub> est indépendant de la valeur absolue de la concentration en DTT et dépend uniquement du rapport DTT<sub>ox</sub>/DTT<sub>red</sub>.

Les valeurs des potentiels redox de différents couples en fonction de leur concentration sont fournies dans la référence ((Gilbert, 1990)).

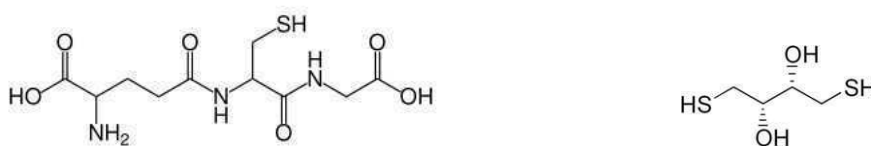


Figure 4 : Formules développées du glutathion réduit (à gauche) et du dithiothréitol réduit (à droite).

La réaction de repliement a donc lieu lorsqu'une protéine réduite est mise en présence d'un agent oxydant. Lors du suivi de la cinétique de repliement oxydant, des prélèvements sont effectués à différents intervalles de temps afin de caractériser l'état du mélange de repliement. Une étape de blocage des formes intermédiaires est nécessaire afin de les stabiliser et de pouvoir ensuite les purifier et les caractériser.

### 3.2.2 Le blocage de la réaction

Deux types de blocage peuvent être effectués :

- Le blocage réversible : dans ce cas on acidifie simplement le mélange de repliement pour atteindre une valeur de pH pour laquelle la formation d'ion thiolate est très défavorisée (typiquement pH 2 à 3). L'intérêt de ce blocage est qu'il est possible de redémarrer le repliement oxydant en remplaçant les molécules dans une solution à pH supérieur à 7. Cette méthode est intéressante pour étudier le repliement d'un intermédiaire de repliement purifié.
- Le blocage irréversible : dans ce cas on ajoute dans la solution un agent ayant la propriété de former une liaison covalente avec les fonctions thiols libres du milieu de repliement (en l'occurrence celles des protéines et celles de l'agent oxydo-réducteur s'il y en a). Certains agents tels le iodoacétamide ou l'acide iodoacétique permettent une alkylation des groupements thiols libres alors que d'autres tels les thiosulfonates (MTSBS : Méthane Thio Sulfonato Butyl Sulfonate) permettent la formation d'une liaison disulfure. Les formules développées de quelques agents bloquants utilisés lors des expériences de repliement oxydant sont présentées sur la Figure 5.

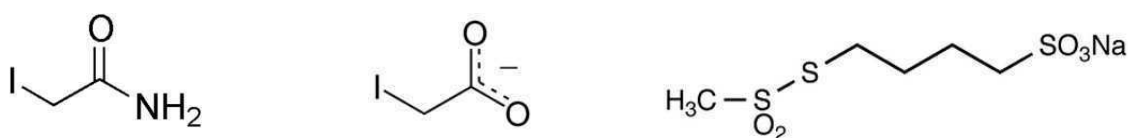


Figure 5 : Formules développées de différents agents bloquant : iodoacétamide (à gauche), iodoacétate (au centre), MTSBS (à droite).

### 3.2.3 La séparation des intermédiaires

La méthode de séparation des intermédiaires est indissociable de la méthode de blocage. Deux types de méthode de séparation sont envisageables :

- Dans le cas où l'objectif est de séparer les espèces en fonction du nombre de cystéines libres, il est préférable d'utiliser un blocage ajoutant une charge à chaque cystéine libre associé à une méthode de purification séparant les espèces en fonction de leur charge ou de leur point isoélectrique (du type chromatographie d'échange d'ions ou isoélectrophorèse).
- Quand l'objectif est de séparer des espèces n'ayant pas de différence au niveau du nombre de cystéines libres mais qui en contrepartie sont partiellement structurées et donc présentent des hydrophobies de surface variables, on utilisera un blocage type acide ou un agent alkylant sans charge du type iodoacétamide, associé à une méthode de purification séparant les espèces en fonction de leur hydrophobie (du type chromatographie phase inverse).

On verra en partie 1.1 des résultats, que ces deux méthodes sont complémentaires, la première permettant d'étudier les phases initiales du repliement car alors, les espèces formées ont des nombres de pont variables et sont peu structurées alors que la deuxième sera préférentiellement utilisée pour étudier les phases finales du repliement lors desquelles on observe des intermédiaires structurés proches de la forme native.

### 3.2.4 La caractérisation des intermédiaires et de leur comportement cinétique

Une fois les intermédiaires bloqués et purifiés, il est possible de les caractériser. Par analyse en spectrométrie de masse, le nombre de ponts formés peut être déterminé. Lorsque les cystéines libres ont été bloquées par un agent alkylant, ces-dernières peuvent être identifiées en réalisant une protéolyse suivie d'une réduction et d'une analyse de masse. Eventuellement le séquençage des peptides obtenus permet de lever les ambiguïtés restantes. Pour obtenir l'identification de l'appariement des ponts, la même analyse est effectuée en supprimant l'étape de réduction. Une fois ces intermédiaires bloqués, s'ils disposent d'une structure, elle peut être caractérisée par les méthodes classiques de biologie structurale comme le dichroïsme circulaire, la cristallographie aux rayons X ou la résonance magnétique nucléaire (RMN).

Lorsqu'un intermédiaire a été bloqué à l'acide, ce blocage est réversible. Il est donc possible d'étudier l'évolution d'un intermédiaire en particulier en l'isolant dans un premier temps en condition acide puis en le plaçant dans un second temps dans un tampon permettant son repliement oxydant.

## 3.3 Classification des mécanismes de repliement des protéines à ponts disulfure

Depuis les travaux d'Anfinsen, le repliement oxydant de nombreuses protéines à ponts disulfure a pu être étudié. Dans leur revue, Arolas *et al.* récapitulent les résultats obtenus et tentent de les comparer afin de classer les mécanismes de repliement des petites protéines à ponts disulfure ((Arolas et al., 2006)). La première limite de cette tentative est liée au fait que les conditions dans lesquelles sont menées les réactions de repliement diffèrent d'une étude à l'autre. Trois paramètres peuvent avoir une influence sur la cinétique de repliement à savoir le pH, la force ionique et l'agent oxydant.

Concernant la force ionique il a été observé que l'ajout de chlorure de sodium dans le milieu de repliement a pour effet d'augmenter la vitesse de repliement ((Chatrenet and Chang, 1993)). Comme on l'a vu en partie 3.1.1 de cette introduction, le pH intervient dans la formation de l'ion thiolate qui est la forme active de la cystéine dans la réaction d'oxydoréduction. Une valeur de pH de 8,5 est considérée comme optimum ((Arolas et al., 2006)). Le choix de l'agent oxydant peut avoir un effet sur la formation de disulfure mixte. D'autre part la présence ou non d'agent réducteur agit sur la vitesse de « reshuffling » (voir partie 3.1.1 de cette introduction). Cependant d'après Arolas *et al.*, pour quatre protéines étudiées dans leur laboratoire (les inhibiteurs de carboxypeptidase de pomme de terre et de sangsue, la protéine anticoagulante de tique et le facteur de croissance épidermique ; en anglais : potato and leech carboxypeptidase inhibitor, tick anticoagulant protein, epidermal growth factor) les différences dans les conditions de repliement n'ont pas d'effet sur la nature des espèces apparaissant et donc pas d'effet sur le mécanisme. ((Chatrenet and Chang, 1993), (Salamanca et al., 2003), (Chang et al., 1994)). Il est donc légitime selon eux de comparer les mécanismes de repliement obtenus dans des conditions qui diffèrent.

Le Tableau 1 récapitule les différents mécanismes de repliement oxydant détaillés à ce jour. Les lettres R et N correspondent à la forme réduite et à la forme oxydée respectivement. La notation XS représente un groupe d'intermédiaires ayant un nombre égal de ponts X mais dont l'appariement peut être hétérogène et en tous cas n'a pas été déterminé. La notation (61-77 ; 73-91) correspond à un intermédiaire ayant deux ponts disulfure dont l'appariement est décrit entre parenthèse. La notation des[30-51] fait référence à un intermédiaire auquel il manque le pont 30-51 mais dont les autres cystéines sont impliquées dans des ponts natifs.

Protéine	Nbre de ponts	Mécanisme proposé	Références
BPTI	3		(Creighton, 1990)  (Weissman and Kim, 1991)
EGF	3	$R \rightleftharpoons 1S \rightleftharpoons \text{des}[6-20] \rightleftharpoons 3S \rightleftharpoons N$	(Chang et al., 2001)
TAP	3	$R \rightleftharpoons 1S \rightleftharpoons 2S \rightleftharpoons 3S \rightleftharpoons N$	(Chang, 1996)
Hirudine	3	$R \rightleftharpoons 1S \rightleftharpoons 2S \rightleftharpoons 3S \rightleftharpoons N$	(Chatrenet and Chang, 1993)
RNAse A	4		(Narayan et al., 2000)
Lysozyme	4		(Chaffotte et al., 1992), (Matagne and Dobson, 1998), (Roux et al., 1999)
LCI	4		(Arolas et al., 2004)
$\alpha$ -lactalbumine	4		(Chang and Li, 2002)
Toxine $\alpha 62$	4		(Ruoppolo et al., 2001)
Cardiotoxine-III	4	$R \rightleftharpoons 1S \rightleftharpoons 2S \rightleftharpoons 3S \rightleftharpoons N$	(Chang et al., 2006)

Tableau 1 : Mécanismes de repliement oxydant décrits à ce jour. Voir le texte pour la signification des abréviations.

A partir de l'analyse de ces différents mécanismes Arolas *et al.* proposent de les classer en trois catégories :

- Les mécanismes de type BPTI : le BPTI (« Bovin Pancreatic Trypsin Inhibitor » en anglais) est la protéine dont le repliement est le mieux connu, notamment grâce aux travaux de Creighton *et al.* ((Creighton, 1990)) et plus tard avec ceux de Weissman et Kim ((Weissman and Kim, 1991)). Ces travaux ont montré que cette protéine à 3 ponts se replie via la formation de 5 populations d'intermédiaires (sur 75 possibles) à 1 et 2 ponts et que toutes ces populations ont des ponts natifs et des éléments de structure natifs (voir partie 3.4 de cette introduction). Les auteurs de cette revue regroupent les repliements de la toxine  $\alpha 62$  (voir la partie 6 de cette introduction pour plus de détail), des cyclotides kalata B1 et McoTI-II (2 protéines de la famille structurale des cyclotides : des petites protéines de 30 acides aminés isolées chez différentes plantes et qui ont la particularité d'avoir un squelette peptidique cyclisé (Daly et al., 2003)) et celui du facteur de croissance de type insuline (insulin-like growth factor : une protéine de 70 résidus proposée comme étant le médiateur de l'hormone de croissance sur les tissus du squelette (Yang et al., 1999)) dans cette classe. Le repliement de ces protéines passe par la formation de formes intermédiaires stables présentant des éléments de structure native et Arolas *et al.* ont rapproché ce type de mécanisme de repliement du modèle « framework ».
- Les mécanismes de type hirudine : le repliement de cette protéine à 3 ponts est bien différent de celui du BPTI, puisqu'on observe dans les premières étapes l'apparition de plus de 30 formes intermédiaires disposant de 1 ou 2 ponts (toujours sur 75 possibles). Au sein de cette population d'intermédiaires, les ponts natifs ne sont pas prédominants. Dans les étapes finales, s'accumulent des isomères mésappariés (« scrambled isomers » en anglais). Ces espèces possèdent un nombre de ponts équivalent à celui de la protéine native mais ne sont pas actives. ((Chatrenet and Chang, 1993)). Les protéines suivantes adoptent ce type de mécanisme : le peptide anticoagulant de tique (en anglais : TAP pour tick anticoagulant peptide (Chang, 1996)), certaines conotoxines ((Fuller et al., 2005)), l'inhibiteur de la carboxypeptidase de pomme de terre (en anglais : potato carboxypeptidase inhibitor (Venhudova et al., 2001)), l'inhibiteur de l' $\alpha$ -amylase d'*Amaranthus* (en anglais :  $\alpha$ -amylase inhibitor (Cemazar et al., 2004)) et la cardiotoxin-III (qui est comme la Toxine  $\alpha 62$  une protéine trois doigts (Chang et al., 2006)). Ce type de repliement s'apparente au modèle de l'effondrement hydrophobe avec un effondrement non-spécifique de la molécule sur elle-même (formation de nombreux intermédiaires à 1 ou 2 ponts non natifs) suivi d'un réarrangement de la structure pour donner la structure native.
- Les mécanismes hybrides : dans ces mécanismes on observe la formation d'un faible nombre d'intermédiaires à 1 ou 2 ponts natifs suivie de la formation d'espèces mésappariées (« scrambled » en anglais) qui se réarrangent en la structure native. L'inhibiteur de carboxypeptidase de sangsue (en anglais : LCI pour Leech Carboxypeptidase Inhibitor) ainsi que le facteur de croissance de l'épiderme (en anglais : EGF pour Epidermal Growth Factor) appartiennent à cette classe ((Arolas et al., 2004), (Chang et al., 2001)). Le repliement oxydant de l' $\alpha$ -lactalbumine est particulier : il s'apparente à celui de l'hirudine en présence de calcium et à celui du BPTI en son absence.

### 3.4 Structures d'intermédiaires de repliement ou de leurs analogues

Un des intérêts majeurs de l'étude du repliement oxydant des protéines à ponts disulfure est qu'il permet l'isolement des espèces intermédiaires. Afin d'étudier leur structure, deux stratégies peuvent être utilisées :

- La première consiste à réaliser le repliement de la molécule puis de bloquer les cystéines libres à l'aide d'un agent alkylant. On étudie ensuite directement les formes intermédiaires purifiées.
- La seconde consiste à concevoir un mutant dont les cystéines correspondant aux cystéines libres dans l'intermédiaire sont mutées en sérines ou en alanines.

La première stratégie est difficile à mettre en œuvre car il est souvent difficile d'obtenir la quantité nécessaire de protéine pour réaliser une étude structurale par purification du mélange d'intermédiaire de repliement. A l'inverse, l'étude à partir des mutants de cystéines est plus indirecte. Il subsiste donc un doute sur le fait que l'intermédiaire et l'analogue obtenu par mutation adoptent précisément la même structure. De plus, le choix de la nature de l'acide aminé de substitution n'est pas neutre. Les premières études de ce type menées par Creighton *et al.* sur le BPTI ont été réalisées par substitution des cystéines libres par des sérines ((van Mierlo et al., 1991b)). Il est considéré que les sérines miment bien les cystéines libres puisqu'elles ne diffèrent des cystéines que par la substitution d'un atome de soufre par un atome d'oxygène. Cependant, l'atome d'oxygène présentant une électronégativité supérieure à celle de l'atome de soufre et la fonction thiol une acidité plus importante que la fonction alcool, la sérine ne remplace pas parfaitement la cystéine réduite. D'autres études ont utilisé une substitution des cystéines par des alanines ((Eigenbrot et al., 1990), (Laity et al., 1997))

#### 3.4.1 Intermédiaires de repliement du BPTI

Le BPTI est également le modèle dont la structure des intermédiaires a été étudiée de la manière la plus exhaustive :

- Lors de son repliement, le BPTI peut emprunter une voie passant par la formation de l'intermédiaire (5-55) qui ensuite donne (5-55 ;14-38). C'est une voie sans issue qui ne permet pas la formation de la protéine native. Le mutant sérine correspondant à l'intermédiaire (5-55) a été étudié en RMN ((van Mierlo et al., 1991b)). Malgré une importante perte de stabilité et une augmentation de la flexibilité de la structure, la conformation de cet analogue d'intermédiaire à 1 pont est semblable à la structure du BPTI natif. Il a été observé que les cystéines 14 et 38 sont bien orientées l'une par rapport à l'autre et qu'elles sont accessibles au solvant. A l'inverse les cystéines 30 et 51 sont enfouies. Cette observation est cohérente avec le fait que (5-55) produit l'intermédiaire (5-55 ; 14-38). La structure tridimensionnelle de C30A/C51A, un analogue de (5-55 ;14-38) obtenue par cristallographie aux rayons X ((Eigenbrot et al., 1990)) montre que les alanines correspondant aux cystéines mutées sont enfouies dans la structure. Cette observation explique pourquoi l'agent oxydant n'a pas accès à ces cystéines et donc l'impossibilité pour cet intermédiaire de permettre la production de la forme native du BPTI. La formation d'un pont reliant l'extrémité N-terminale à l'extrémité C-terminale diminue suffisamment l'espace conformationnel pour que l'ensemble de la structure native apparaisse.

- L'intermédiaire (30-51) est un des intermédiaires à 1 pont qui permettent la production de l'intermédiaire des [14-38], lui-même constituant l'unique intermédiaire à deux ponts productif du mécanisme. La structure tridimensionnelle d'un analogue de cet intermédiaire obtenue par RMN ((van Mierlo et al., 1993)) montre que l'organisation générale de la molécule est comparable à celle de la molécule native bien que certaines parties ne soient pas structurées. En effet les 15 résidus N-terminaux ainsi qu'une boucle englobant les résidus 37 à 40 sont beaucoup plus mobiles et ne forment pas les contacts nOe attendus. A partir de cet intermédiaire se forme l'intermédiaire (14-38 ; 30-51) par fermeture du pont 14-38. Le mécanisme proposé dans le Tableau 1 indique que l'intermédiaire (14-38 ; 30-51) ne peut pas produire directement la forme native ; le pont natif 14-38 doit préalablement être réduit pour que les ponts non natifs 5-14 ou 5-38 puissent se former. Ces deux intermédiaires donnent (5-55 ; 30-51) qui lui est productif. L'étude en R.M.N de (14-38 ; 30-51) ((van Mierlo et al., 1991a)) montre que cet intermédiaire est proche de la structure native même si les extrémités N et C-terminales sont partiellement déstructurées. Une nouvelle fois, un intermédiaire disposant d'une structure native et dont il ne manque qu'un seul pont représente une voie sans issue. L'accessibilité des cystéines ne semble pas ici être en cause puisque lors du repliement du BPTI les deux formes à disulfure mixte de l'intermédiaire (14-38 ; 30-51) ont pu être isolées. L'étape de formation du disulfure mixte n'est donc pas limitante dans la réaction d'oxydation des cystéines 5 et 55. C'est donc la seconde étape consistant en l'attaque du disulfure mixte par l'ion thiolate de la cystéine libre restante qui semble défavorisée (Figure 3). Les auteurs proposent deux hypothèses pour expliquer ce phénomène. La première implique l'absence de structure native au niveau des extrémités N et C terminales de l'intermédiaire (14-38 ; 30-51) qui contiennent les résidus 5 et 55 empêche l'apparition d'une géométrie entre les cystéines libres compatible avec la formation du dernier pont (hypothèse entropique). La seconde postule la nécessité d'existence d'une conformation de haute énergie (état contraint ou « frustré ») pour permettre l'attaque du disulfure mixte par l'ion thiolate qui empêcherait la fermeture du pont.

Les résultats précédents montrent que les intermédiaires (5-55) et (14-38 ; 30-51) ont des structures natives. Cela implique que les ponts disulfure du BPTI ne constituent pas des éléments qui définissent la topologie adoptée par cette protéine. Ils ne sont manifestement pas nécessaires à l'acquisition de la structure native. Cependant leur contribution à la stabilité est manifeste.

### **3.4.2 Intermédiaires de repliement de la RNase A :**

La RNase A est une protéine à 4 ponts qui se replie via deux intermédiaires à 3 ponts natifs tous deux capables de produire la forme native. Ces intermédiaires sont des[40,95] et des[65-72]. Des analogues de des[40-95] et des[65-72], obtenus par la substitution des cystéines libres par des résidus sérine dans des[40-95] et alanine dans des[65-72], ont été étudiés par RMN ((Shimotakahara et al., 1997), (Laity et al., 1997)). Ces études montrent que ces deux analogues ont des structures proches de la structure native et que tous deux sont moins stables que la protéine native. Cette perte de stabilité est imputée à une déstabilisation de leur cœur hydrophobe.

### 3.4.3 Intermédiaires de repliement du lysozyme du blanc d'œuf de poule:

Lors de son repliement, le lysozyme forme trois intermédiaires à trois ponts natifs à savoir des[6-127], des[64-80] et des[76-94]. L'intermédiaire des[76-94] est le plus abondant. Sa conversion en protéine native est beaucoup plus lente que celle de des[6-127] ((van den Berg et al., 1999)). L'étude RMN directe de des[76-94] bloqué au iodoacétamide montre que sa structure est comparable à celle de la protéine native ((van den Berg et al., 1999)). Un analogue de cet intermédiaire dans le lysozyme humain dont les cystéines libres étaient mutées en alanine a également été étudié par cristallographie aux rayons X ((Inaka et al., 1991)). La structure de cet analogue montre que l'alanine 77 (position équivalente à la cystéine 76 dans le lysozyme du poulet) est accessible au solvant alors que l'alanine 94 est enfouie. En considérant que l'analogue de la protéine humaine a une structure équivalente à celle de l'intermédiaire de la protéine de poulet, ces résultats expliquent l'impossibilité du pont 76-94 à se former par l'inaccessibilité de la cystéine 94.

### 3.4.4 Intermédiaire de repliement du LCI:

L'intermédiaire III-A identifié lors du repliement du LCI comme étant l'intermédiaire des[22-58] a été étudié en RMN ((Arolas et al., 2004), (Arolas et al., 2005)). Exceptionnellement, la molécule utilisée n'était ni un analogue, ni un intermédiaire alkylé mais cet intermédiaire a été purifié en conditions acides. Les expériences de RMN ont été menées à pH 3,5 afin d'empêcher l'oxydation des cystéines libres. Cet intermédiaire constitue le principal piège cinétique lors du repliement du LCI. La structure de III-A s'apparente à la structure native. Une nouvelle fois l'hypothèse proposée pour expliquer la difficulté à fermer le dernier pont implique un enfouissement des cystéines libres.

### 3.4.5 En résumé :

Toutes les structures d'intermédiaires de repliement étudiées jusqu'à aujourd'hui montrent que ces espèces adoptent des structures très proches de la forme native. L'étude structurale d'espèces dans lesquelles il manque deux ponts natifs a été menée uniquement pour le modèle BPTI. Ces espèces ont déjà acquis la topologie de la structure native et même une bonne partie de la structure secondaire, seules certaines parties étant déstructurées. Dans le cas du BPTI, il semble qu'aucun des ponts disulfure n'est nécessaire à l'acquisition de l'agencement natif.

Les intermédiaires dont il manque uniquement un pont natif ont généralement une structure de type natif. Ces études structurales ont été le plus souvent mises en œuvre afin d'expliquer l'incapacité de ces formes à produire la protéine oxydée. Dans la grande majorité des cas l'existence d'une structure de type natif est associée à l'enfouissement des cystéines libres ce qui explique l'incapacité pour ces cystéines d'avoir accès à l'agent oxydant. Ce type d'intermédiaire a été baptisé par Scheraga *et al.* ((Narayan et al., 2000)) intermédiaire « locking-in » puisque les cystéines libres se trouvent être piégées dans leur structure. L'intermédiaire (14-38 ; 30-51) du BPTI est à ce jour le seul intermédiaire de protéine à ponts disulfure qui n'est pas de type « locking-in » puisqu'on retrouve des formes contenant des disulfure mixte de (14-38 ; 30-51) lors du repliement. Les cystéines libres se trouvent à proximité puisque la structure est de type natif, elles sont accessibles au solvant et pourtant le pont ne se forme pas. Ceci implique que dans certains cas, l'étape limitante est l'attaque du disulfure mixte par l'ion thiolate de la cystéine libre. Les hypothèses entropiques ou de frustration ont toutes deux été proposées pour expliquer l'impossibilité de voir cet intermédiaire produire directement la forme native (voir partie 3.1.2 de cette introduction).



---

## 4 LE MODELE D'ETUDE : LA TOXINE $\alpha$ DE NAJA NIGRICOLLIS, UNE PROTEINE A TROIS DOIGTS

---

### 4.1 La toxine $\alpha$

La toxine  $\alpha$  est une neurotoxine à trois doigts, isolée à partir du venin d'un cobra, *Naja nigricollis* ou cobra cracheur noir (Figure 6)



Figures 6 : *Naja nigricollis* ou cobra cracheur noir.

Ces cobras sont très répandus en Afrique occidentale et jusqu'en Afrique du sud. Comme les autres serpents cracheurs, ils peuvent utiliser leur venin en mordant ou en projetant leur venin jusqu'à deux ou trois mètres. L'envenimation par ce cobra est essentiellement cardiotoxique, elle conduit à des arythmies et des vasoconstrictions puis dans un deuxième temps à des nécroses. Son action neurotoxique est moins prononcée que pour d'autres espèces du genre *Naja*.

La toxine  $\alpha$  est une protéine constituée de 61 résidus dont la séquence est présentée sur la Figure 7.

Cette protéine contient 4 ponts disulfure, sa structure secondaire est composée de 5 brins  $\beta$  organisés en deux feuillets antiparallèles et de 5 coudes  $\beta$  (Figure 8). Cette protéine appartient à la famille structurale des protéines trois doigts. Dans la base SCOP ((Murzin et al., 1995)), cette famille correspond au domaine « Snake Toxin-like ».

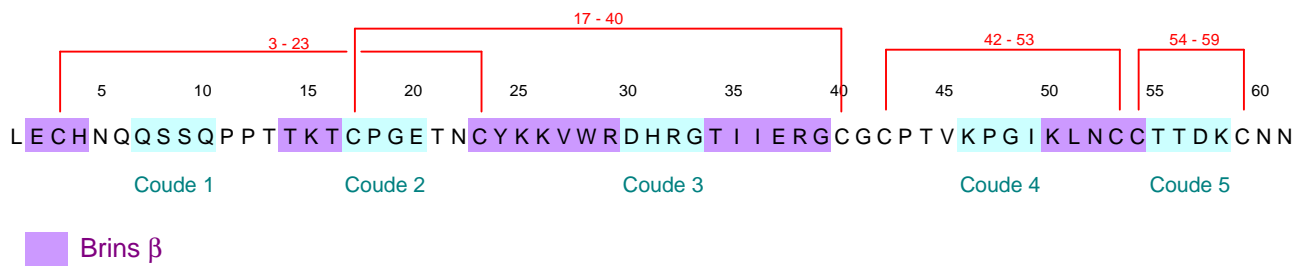
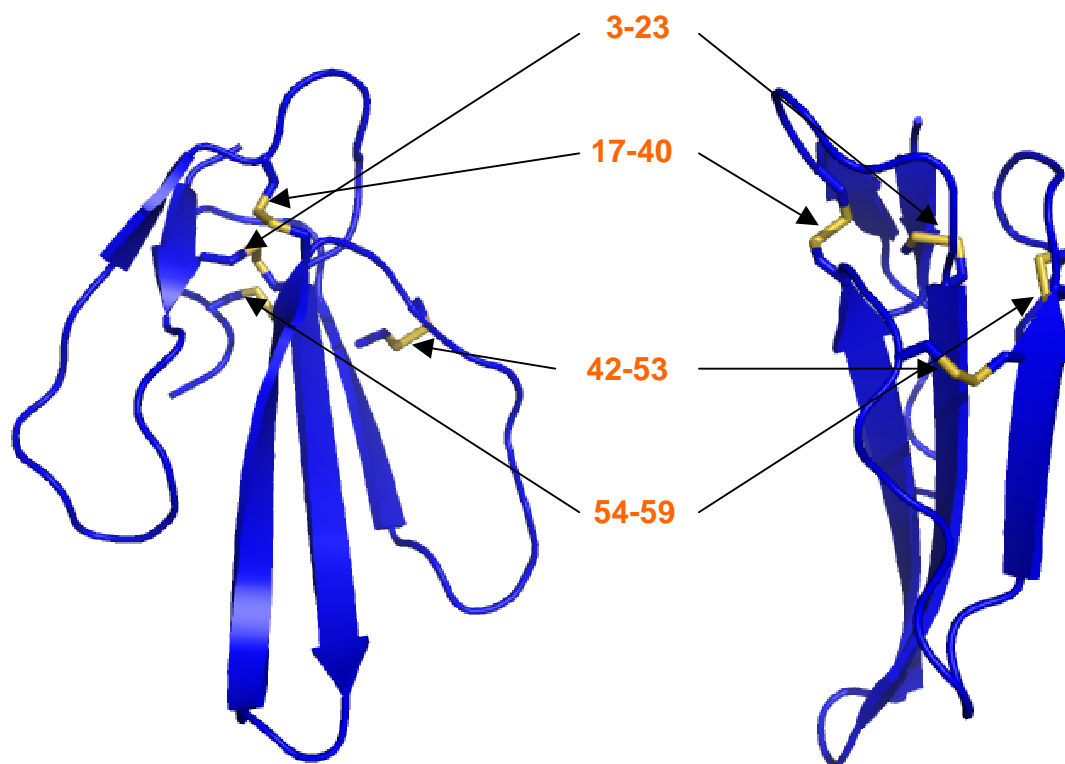


Figure 7 : Séquence, appariement des ponts disulfure et éléments de structure secondaire de la toxine  $\alpha$



Figures 8: Structure cristallographique (code PDB : 1IQ9) et ponts disulfures de la toxine  $\alpha$

## 4.2 Fonctions exercées par les toxines à trois doigts

Les toxines à trois doigts isolées dans les venins de serpent présentent une certaine diversité de modes d'action toxiques. L'article présenté en annexe 1 présente les fonctions exercées par les toxines à trois doigts en détail.

### 4.2.1 Inhibiteurs de récepteurs impliqués dans le fonctionnement du système nerveux :

Certaines neurotoxines à trois doigts (dont fait partie la toxine  $\alpha$ ) bloquent la communication nerveuse en se liant sur les récepteurs à l'acétylcholine ((Servent, 2001); (Nirthanan and Gwee, 2004)). Ces toxines sont exclusivement des antagonistes compétitifs de l'acétylcholine.

Cette classe est elle-même découpée en quatre groupes en fonction du sous-type de récepteur ciblé :

- $\alpha$ -neurotoxines : leur chaîne polypeptidique peut être courte (60-62 résidus) ou longue (69 résidus) et elles contiennent quatre ponts disulfure. Ces toxines ciblent exclusivement les récepteurs musculaires à l'acétylcholine.
- $\kappa$ -neurotoxines : elles sont longues (66 résidus), contiennent 5 ponts disulfure avec un pont additionnel dans la boucle II et forment des dimères. Elles ciblent exclusivement les récepteurs neuronaux à l'acétylcholine.
- $\alpha/\kappa$ -neurotoxines : leur chaîne est longue (66 à 74 résidus), elles contiennent comme les  $\kappa$ -neurotoxines un pont additionnel dans la boucle II mais se lient à la fois aux récepteurs musculaires et aux récepteurs neuronaux.
- Neurotoxines non-conventionnelles : leur chaîne est longue, elles ont un pont additionnel dans la boucle I, pour la plupart d'entre elles, aucune cible n'a pu être identifiée.

D'autres neurotoxines ciblent des récepteurs couplés à la protéine G tels que les récepteurs muscariniques ou adrénergiques ((Ellis, 1997); (Christopoulos and Kenakin, 2002)). On trouve les récepteurs muscariniques au niveau du système nerveux et l'activation de ces récepteurs conduit à la modulation de l'activité de différentes molécules de signalisation comme des enzymes (phospholipase, adénylate cyclase, MAP kinase) ou des canaux ioniques (canaux calciques voltage-dépendant ou GIRK). A la différence des toxines nicotiques qui sont toutes des inhibiteurs compétitifs, les toxines muscariniques peuvent agir soit comme antagonistes, soit comme modulateurs allostériques. Les récepteurs adrénergiques sont quant à eux potentiellement exprimés dans toutes les cellules. Au niveau du système nerveux central, ils sont impliqués dans les processus cognitifs et d'apprentissage. Les toxines adrénergiques se lient à ce type de récepteurs.

#### **4.2.2 Inhibiteurs de protéines membranaires non neuronales**

La calciseptine et son homologue la FS2 inhibent les canaux calciques de type L, provoquant le relâchement des muscles lisses et inhibant la contraction cardiaque ((Dewille et al., 1991); (Yasuda et al., 1994)). La cardiotoxine A5 et la dendroaspine se lient aux intégrines selon des modes de liaison qui diffèrent mais qui dans les deux cas conduisent à l'inhibition de l'agrégation plaquettaire ((Rahman et al., 2000))

#### **4.2.3 Inhibiteurs d'enzymes**

Les fasciculines se lient à l'acétylcholinestérase et l'inhibent de façon non-compétitive en bloquant la voie d'accès du substrat au site catalytique ((Karlsson et al., 1984)). Les hemextines A et B, lorsqu'elles s'associent en un complexe AB, inhibent l'activité protéasique du facteur VIIa ce qui empêche la coagulation ((Banerjee et al., 2005)).

#### **4.2.4 Cytotoxines**

Il a été montré que leur toxicité s'exerce par la dépolarisation des cardiomyocytes ce qui conduit à un arrêt cardiaque de la victime ((Kumar et al., 1997)). Le mécanisme proposé pour expliquer ces effets implique la liaison de ces toxines à la membrane des cardiomyocytes, puis une oligomérisation conduisant à la formation de pores aboutissant à la fuite des solutés contenus dans le cytoplasme.

## 5 LA FAMILLE STRUCTURALE DES PROTEINES A TROIS DOIGTS

Cette partie est un résumé en français de l'extrait d'ouvrage présenté en annexe 1.

### 5.1 Le domaine à trois doigts

D'après la base SCOP ((Murzin et al., 1995)), l'unité de base de classification des protéines est le domaine. Dans les protéines naturelles, ces domaines peuvent soit être isolés et constituer par eux-même une protéine monodomaine, soit être associés à d'autres domaines pour former des protéines multidomaines ((Lo Conte et al., 2002)). Dans un article paru en 2006 Cheek *et al.* ((Cheek et al., 2006)) ont proposé le regroupement des 2945 domaines de petites protéines riches en pont disulfure en 41 classes en fonction de leur similarité de structure. Selon cette classification, une petite protéine riche en ponts disulfure type est composée de  $57 \pm 27$  résidus, contient  $3 \pm 1$  ponts disulfure. La plupart de ces domaines proviennent d'organismes eucaryotes. On retrouve ces domaines dans différents venins où ils exercent une grande diversité de fonctions toxiques ((Cheek et al., 2006)). La Figure 9 A présente la structure cristallographique de la toxine  $\alpha$  de *Naja nigricollis*, typique du domaine à trois doigts ((Gilquin et al., 2003)). La topologie de cette protéine forme trois boucles (ou doigts) qui émergent d'un cœur hydrophobe situé dans la partie apicale (Figure 9). La structure secondaire est composée de 5 brins  $\beta$  s'organisant en deux feuillets. Des protéines adoptant ce repliement ont été initialement découvertes dans les venins de serpent ((Low et al., 1976), (Tsernoglou and Petsko, 1976)). Depuis lors, ce motif structural a été retrouvé dans de nombreuses protéines eucaryotes animales telles que les domaines extracellulaires de divers récepteurs comme le récepteur à l'activine de type II (en anglais, Activin receptor type II), le récepteur de la protéine de la morphogénèse osseuse (BMP receptor), le récepteur du facteur de croissance transformant bêta de type II (TGF beta type II receptor), le récepteur de l'activateur du plasminogène de type urokinase (UPAR), le récepteur ly-6 et le récepteur CD59 ((Tsetlin, 1999)). La protéine LYNX qui est associée à la membrane ((Miwa et al., 2006)) ainsi que la protéine soluble SLURP ((Adermann et al., 1999)) ont également été classées dans la famille des protéines trois doigts.

### 5.2 Classification du domaine à trois doigts au sein du groupe des petites protéines riches en ponts disulfure

La classification des petites protéines riches en ponts disulfure est basée sur leurs ponts : soit selon leur positionnement dans l'espace ((Harrison and Sternberg, 1996), (Mas et al., 2001), (Cheek et al., 2006)), soit selon leur appariement dans la séquence ((Gupta et al., 2004), (Chuang et al., 2003)).

En analysant les structures disponibles des petites protéines à ponts disulfures, Harisson et Sternberg ont identifié un motif structural qu'ils ont baptisé « disulfide  $\beta$ -cross » ((Harrison and Sternberg, 1996)) (Figure 10). Ce motif est composé de deux cystines dont chacune des cystéines C-terminales appartient à un brin  $\beta$  d'un même «  $\beta$ -hairpin ». Dans ce motif, l'appariement des ponts est croisé dans la séquence (Figure 9 C). Les auteurs ont défini la chiralité de cette structure de la manière suivante : dans certaines familles de protéines, les segments contenant les cystéines Cys2N et Cys2C forment une hélice avec pas à droite en passant par Cys1C (Figure 10 A) ; dans le cas des protéines trois doigts, ces mêmes segments forment une hélice à pas à gauche (Figure 10 B).

Ils ont souligné l'agencement particulier des ponts 1, 2 et 4 qui sont regroupés dans l'espace (Figure 9 B), on parle alors de « cystine stacking ». Selon ces auteurs, les protéines trois doigts sont définies de la manière suivante : ce sont des protéines composées d'un « disulfide  $\beta$ -cross » et d'un « cystine stack » impliquant trois ponts.

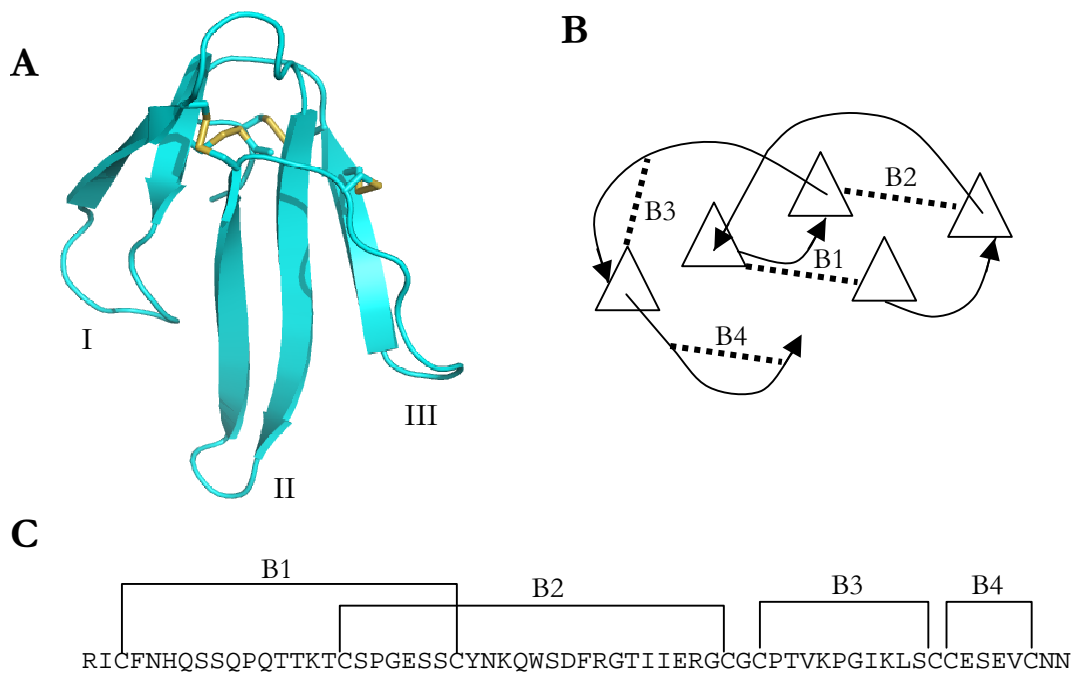


Figure 9 : A : Structure cristallographique d'une protéine typique du repliement trois doigts. B : Organisation des structures secondaires et connectivité du repliement trois doigts. Notons ici que B1, B2, B4 sont parallèles alors que B3 adopte une orientation différente. C : Séquence en acides aminés et appariement des ponts de l'érabutoxine b.

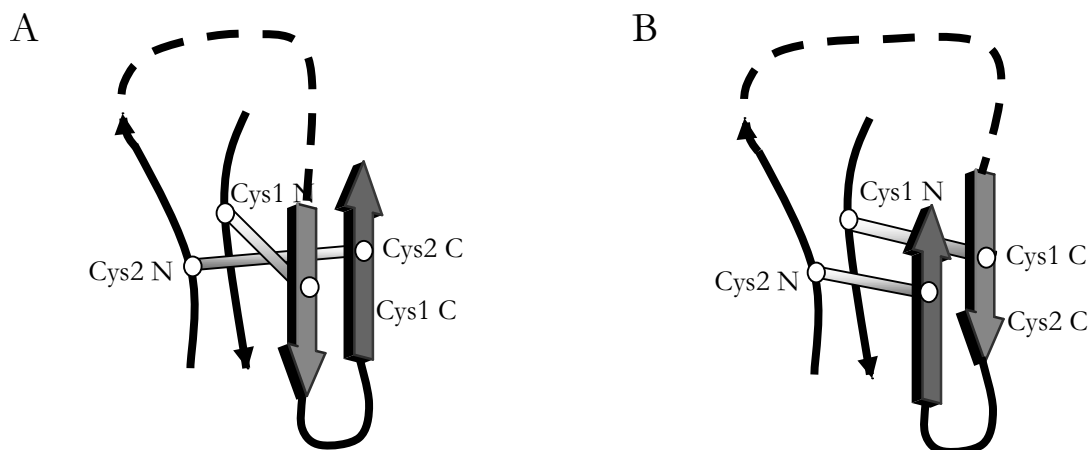


Figure 10: A : Le « disulfide  $\beta$ -cross » comme définit par Harrison et Sternberg ((Harrison and Sternberg, 1996)). B : Quand les éléments du «  $\beta$ -cross » ont une connectivité gauche, le croisement spatial des ponts disparaît, ils sont alors parallèles.

Mas *et al.* ont développé une méthode automatique, appelée KNOT-MATCH afin de classer les protéines à pont disulfure en fonction de la topologie de leurs ponts. Le programme calcule un RMSD (Root Mean Square Distance) sur différentes combinaisons de ponts disulfure puis classe les protéines en fonction de cette mesure ((Mas et al., 2001)). Cette méthode a généré 13 classes de protéines. Dans les classes contenant les protéines trois doigts, trois ponts disulfure étaient systématiquement agencés comme les barreaux d'une échelle avec un RMSD inférieur à 2,5 Å. Cet agencement rappelle le « cysteine stacking » de Harrison et Sternberg. Cependant, tous les éléments de cette classe ne sont pas des protéines trois doigts. On trouve également des domaines dont la topologie est très différente tels que le « Nerve Growth Factor » ou les « Defensin ». Par conséquent, la méthode KNOT-MATCH n'est pas capable à elle seule de définir les limites du domaine trois doigts. Plus récemment Cheek *et al.* ont classé les 2945 petites protéines à pont disulfure apparaissant dans la banque de donnée PDB en 41 groupes ((Cheek et al., 2006)). L'architecture des éléments de structure secondaire et la topologie furent utilisés comme critères de classification. Dans le cadre de cette classification, les protéines trois doigts furent classées dans un groupe appelé « Knottin-like II », l'expression « knottin-like » faisant directement référence au disulfide  $\beta$ -cross de Harrison et Sternberg. Le groupe « knottin like » dans son ensemble représente 40% des structures des petites protéines à ponts disulfure. Il a été découpé en trois sous-groupes en fonction de la connectivité des éléments structuraux : dans le groupe « knottin-like II », les quatre demi-cystines (formant le « disulfide  $\beta$ -cross » de Harrison et Sternberg) sont localisées sur quatre éléments structuraux dont la connectivité est gauche. Cet agencement implique que le croisement des deux premiers ponts n'existe que dans la séquence (Figure 9 C) et non dans l'espace comme cela est représenté sur la Figure 10 A. Cependant le groupe des knottin-like II n'est pas uniquement composé de protéines à trois doigts mais regroupe également des protéines dont la topologie diffère comme les protéines hirudin-like ou le domaine N-terminal du « granulin-like repeat ». Par conséquent, les critères basés sur les ponts disulfure utilisés par Cheek *et al.* pour classer les petites protéines riches en ponts disulfure ne reflètent pas strictement l'organisation topologique des protéines à trois doigts. Dans le passé, les protéines à trois doigts ont été rapprochées d'une famille structurale adoptant une topologie de type « plant lectin like » (comme l'agglutinine de germe de blé par exemple) pour définir un repliement « toxin-agglutinin » ((Drenth et al., 1980)).

Toujours pour classer les protéines à ponts disulfure, la méthode de Chuang *et al.* s'appuie sur la connectivité des ponts dans la séquence. Une bonne corrélation entre la structure de ces protéines et leur profil d'appariement a été obtenu ((Chuang et al., 2003)) puis amélioré en ajoutant l'espace entre les cystéines comme paramètre supplémentaire ((Gupta et al., 2004)). Le nombre de ponts, leur connectivité et leur espacement ont été convertis en une série de nombres appelés « disulfide signature ». Les nombres ayant un rang impair dans la signature représentent le nombre de résidus compris entre les deux cystéines d'un pont; les nombres ayant un rang pair indiquent le nombre de résidus séparant les deux premières cystéines de deux ponts consécutifs dans la séquence (représentés en italique dans l'exemple ci-après). Par exemple la « disulfide signature » de l'erabutoxine a (Figure 9 C) est 21-14-24-26-11-12-5. Des distances mathématiques peuvent être aisément calculées à partir de ces signatures, permettant la comparaison et donc la classification des protéines à ponts disulfure. Quand on tient compte de cinq ponts disulfure, cette méthode est suffisamment efficace pour séparer et regrouper toutes les protéines trois doigts en un seul groupe. Mais lorsqu'on s'intéresse aux protéines à quatre ponts disulfure, le groupe des protéines trois doigts est contaminé par des protéines de la famille des protéines « Laminin EGF-like ». Ainsi, même si le profil d'appariement semble constituer le caractère le plus efficace pour classer les protéines à ponts disulfure, il est insuffisant pour définir la topologie à trois doigts.

## 5.3 Déterminants structuraux conservés au sein de la famille des protéines trois doigts

Dans ce paragraphe, sont résumés les résultats principaux de l'étude menée au sein de notre laboratoire concernant la définition du domaine à trois doigts ((Galat et al., 2008) fourni en annexe 2). Le parti a été pris d'étudier uniquement les protéines dont la structure a été résolue et qui adoptent un repliement à trois doigts. Les données structurales de 34 toxines et de 16 domaines extracellulaires de récepteurs ont été analysées (Table 1 de l'article fourni en annexe 2).

Une terminologie spécifique aux protéines trois doigts a été définie pour désigner les différents éléments de structure des protéines trois doigts. La Figure 11 présente la structure de la toxine  $\alpha$  de *Naja nigricollis* ainsi que cette terminologie.

### 5.3.1 Analyse des séquences

En utilisant la séquence de la toxine  $\alpha$  de *Naja nigricollis* comme référence (code PDB : 1IQ9), les scores de similitude ont été calculés deux à deux pour toutes les séquences étudiées. Ce score variait entre 86 et 30% pour les séquences des toxines et était inférieur à 25% pour certains domaines extracellulaires de récepteur. Cette différence provient pour l'essentiel de boucles plus longues et de nombreuses mutations dans les doigts des domaines extracellulaires de récepteur. Les 6 demi-cystines des ponts 1, 2 et 4 et une asparagine adjacente à la demi-cystine C-terminale du pont 4 sont strictement conservées à travers la famille des protéines à trois doigts (Figure 12). Les demi-cystines du pont 3 sont hautement conservées mais pas strictement car il existe une exception dans le premier domaine de uPAR (1YWH1). Les cystéines C-terminales du pont 1 et du pont 2 ont toutes deux un résidu adjacent hautement conservé : un résidu hydrophobe (souvent aromatique) dans le premier cas et une glycine dans le second. Cette-dernière est strictement conservée au sein des toxines. D'autres éléments de séquence peuvent varier de manière substantielle au sein de la famille des protéines trois doigts. Notamment, la longueur et la composition des doigts, la présence d'hélices  $\alpha$  ou de ponts disulfure supplémentaires (Figure 1 de l'annexe 2).



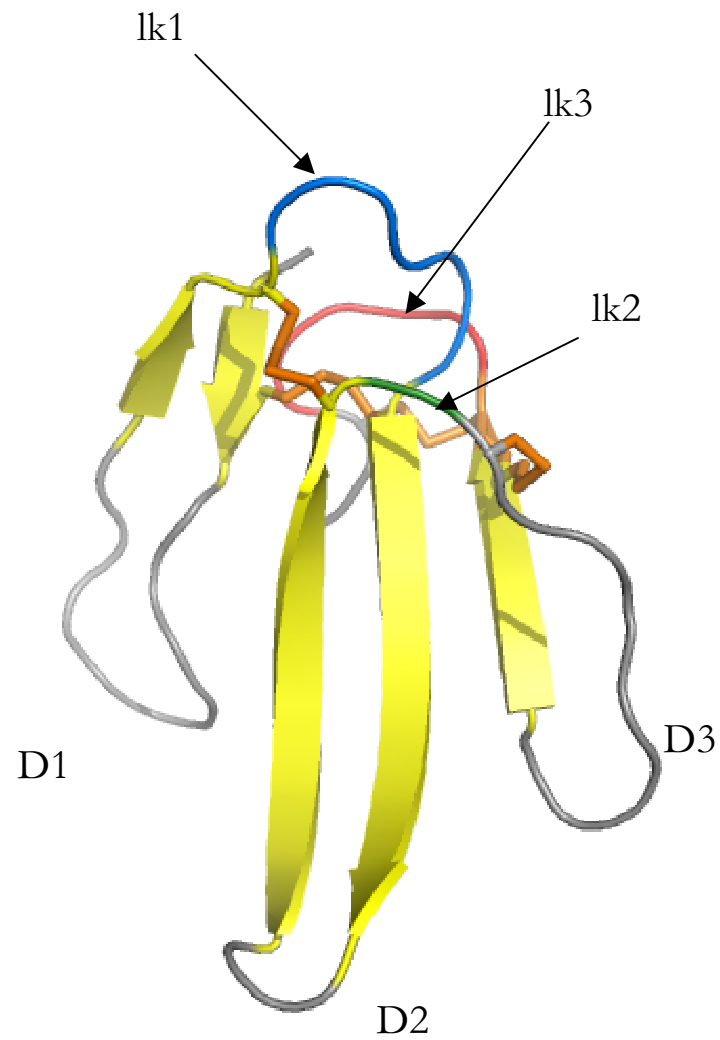


Figure 11 : Structure de la toxine  $\alpha$  de *Naja nigricollis* et éléments de structure. lk1 à 3 : linker 1 à 3 ; D1 à 3 : doigt 1 à 3

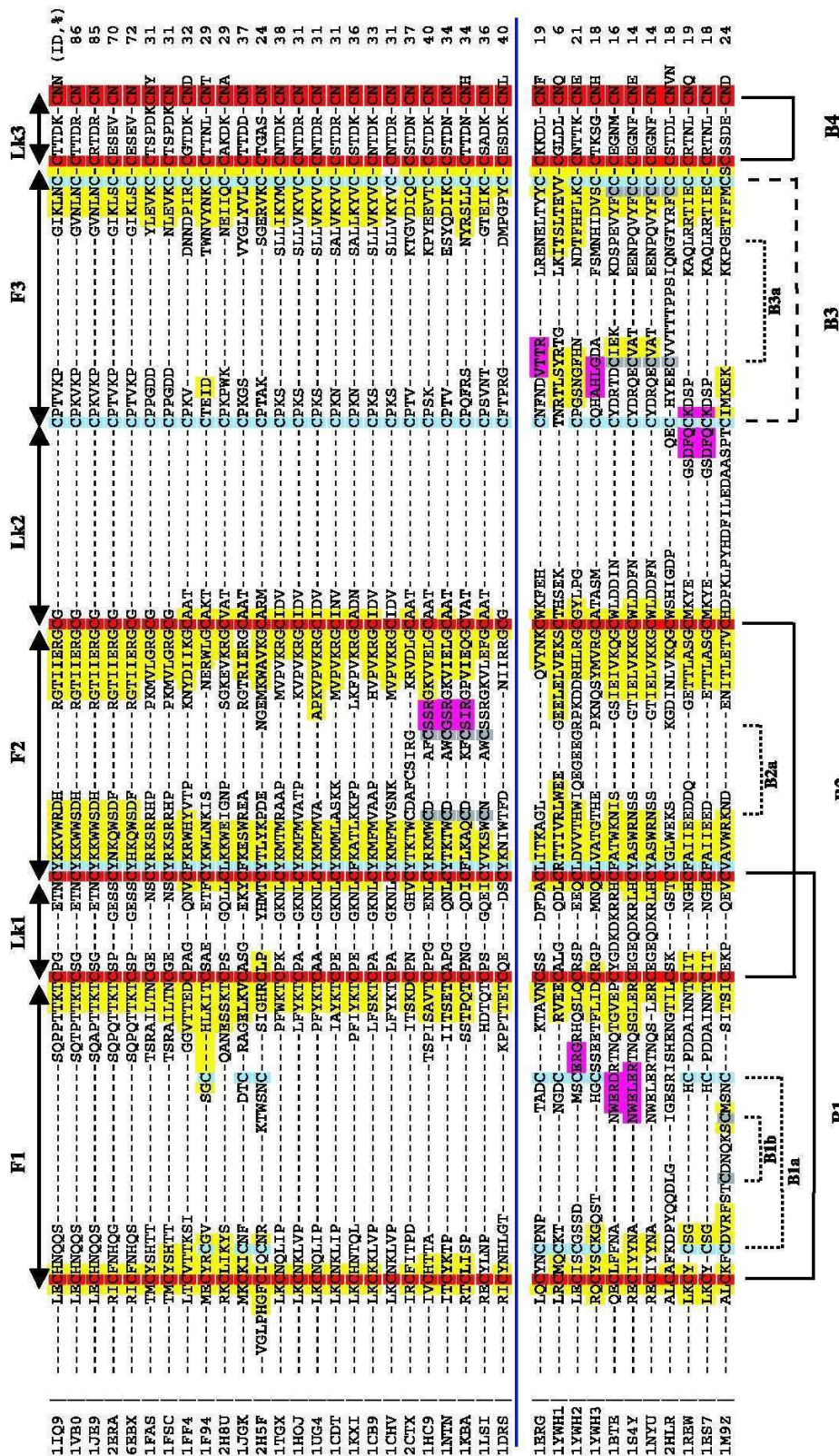


Figure 12 : Alignement des séquences des protéines trois doigts étudiées dans l'article ((Galat et al., 2008)). Les hélices  $\alpha$  et les feuillets  $\beta$  sont surlignés en magenta et en jaune respectivement. Les résidus strictement conservés sont surlignés en rouge, les cystéines hautement conservées en cyan et celles spécifiques de certaines sous-classes en gris.

### 5.3.2 Analyse des structures :

En analysant l'alignement des séquences des protéines trois doigts, il apparaît que 7 résidus sont parfaitement conservés à travers cette famille : les ponts 1, 2, 4 et l'asparagine C-terminale. La suite de cette analyse a consisté à déterminer si la conservation stricte de ces résidus dans la séquence se traduisait par une conservation de certains motifs structuraux.

#### 5.3.2.1 le « cluster » de cystéines :

Les ponts 1, 2 et 4 s'organisent en un cluster. Leur agencement spatial rappelle les barreaux d'une échelle (Figure 13) comme cela avait mis en évidence par le logiciel KNOT-MATCH de Mas *et al.*.

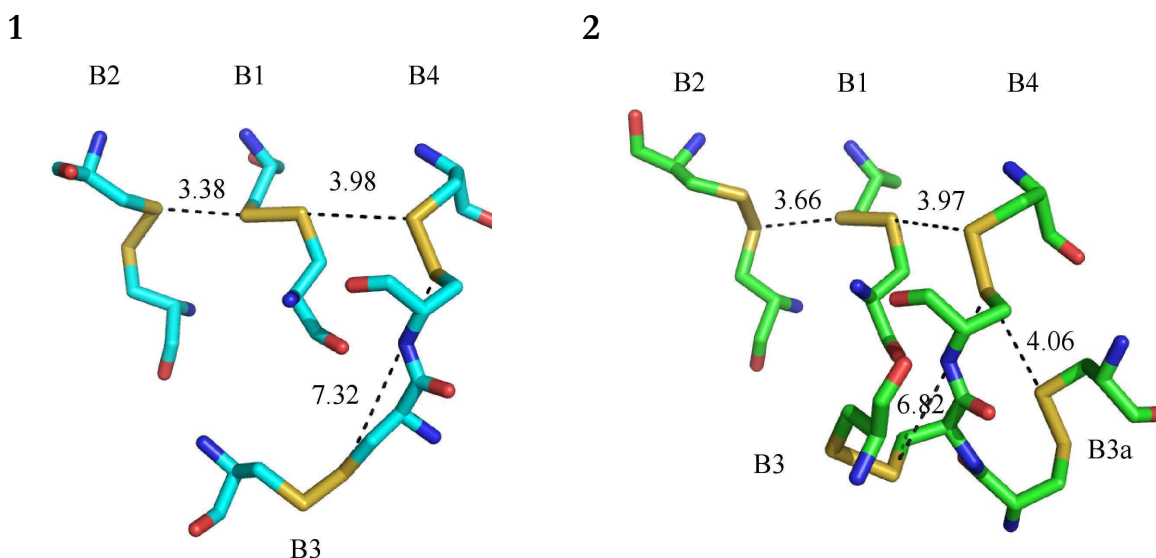


Figure 13 : Cluster de cystéines pour deux protéines trois doigts. 1 : la toxine  $\alpha$  de *Naja nigricollis* (1IQ9), 2 : le domaine extracellulaire du récepteur de l'activine type IIB murin (1S4Y).

La variabilité dans le positionnement des ponts 1,2 et 4 impliqués dans le « cluster » et l'effet induit par l'ajout du pont 3 sur cette mesure ont été déterminés. Pour ce faire, les trois ponts de la toxine  $\alpha$  de *Naja nigricollis* (1IQ9), en considérant cette protéine comme une référence, ont été superposés sur les cystines de chacune des autres protéines trois doigts de l'étude. Comme on peut le voir sur la Figure 14 (barres noires), la plupart des valeurs de RMSD varient entre 0,5 et 1 Å, avec une grande majorité de valeurs proches de 0,5Å. Cette valeur atteint 1,5Å dans quatre cas qui correspondent à des structures de complexes. L'organisation spatiale du cluster de cystéine est donc parfaitement conservée au sein des protéines trois doigts. Les résultats pour le pont 3 sont bien différents (Figure 14, barres blanches). En effet, pour les autres neurotoxines courtes se liant aux récepteurs à l'acétylcholine, le RMSD était inférieur à 0,7Å. Pour les toxines adrénérgiques, muscariniques ainsi que les neurotoxines longues, le RMSD atteint 1 Å. Enfin il varie entre 1,5 et 3,9 Å dans le cas des domaines extracellulaires de récepteurs. Lorsque l'on tient compte de ce pont, la valeur du RMSD semble être un indicateur de la proximité fonctionnelle avec la toxine  $\alpha$ . Les différentes familles fonctionnelles de protéines trois doigts semblent se différencier par le positionnement spatial du pont trois.

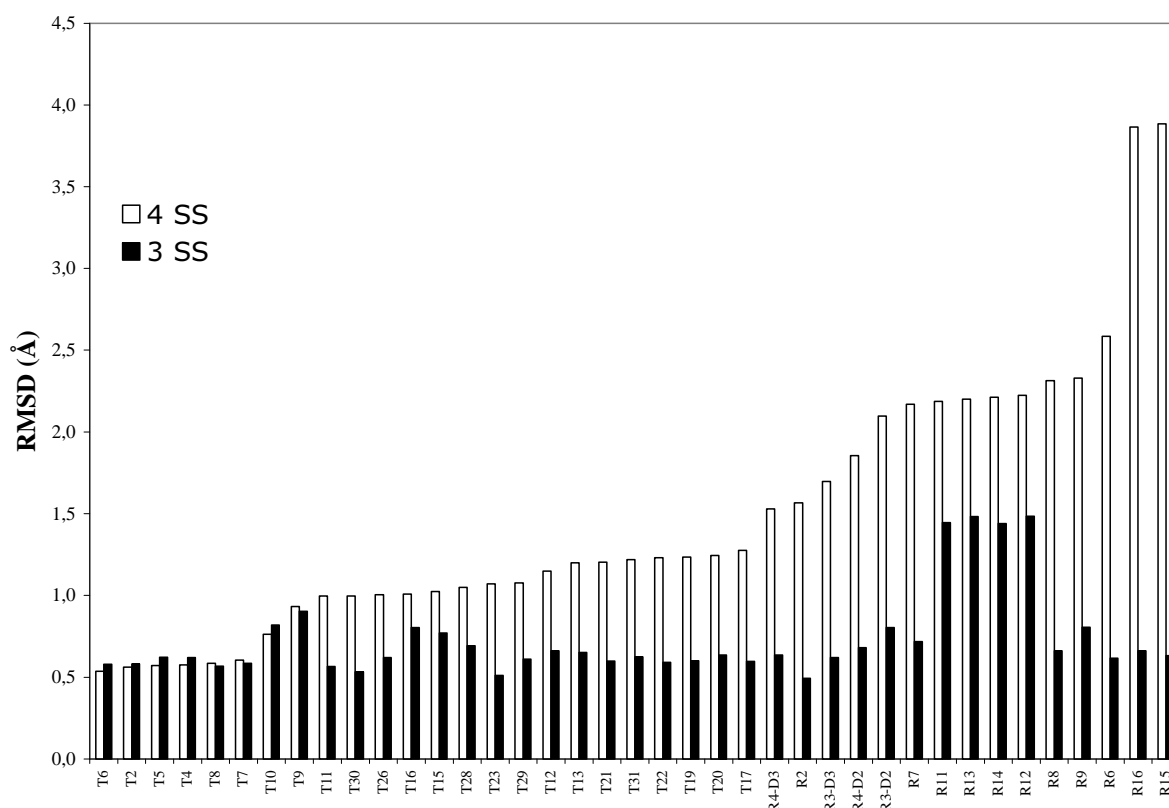


Figure 14 : Valeurs de RMSD du réseau de cystines calculées entre la toxine  $\alpha$  (code PDB : 1IQ9) utilisée comme référence et chacune des autres protéines trois doigts. Les barres noires correspondent aux ponts 1, 2 et 4 et les barres blanches en ajoutant le pont 3. Les protéines sont rangées en par ordre croissant du RMSD obtenu avec les quatre ponts. Les abscisses correspondent aux codes donnés dans la table 1 de l'article fourni en annexe 2.

### 5.3.2.2 L'asparagine C-terminale conservée :

L'analyse des interactions existant dans les structures de plusieurs protéines trois doigts a montré que l'asparagine C-terminale conservée est impliquée dans de nombreuses interactions. Dans toutes les structures étudiées, elle pointe vers l'intérieur de la molécule. La chaîne latérale est systématiquement impliquée dans des liaisons hydrogènes avec le squelette peptidique de deux résidus adjacents aux cystéines du pont 1 (Figure 15). Ce réseau d'interactions est remarquablement conservé dans la famille des protéines trois doigts.

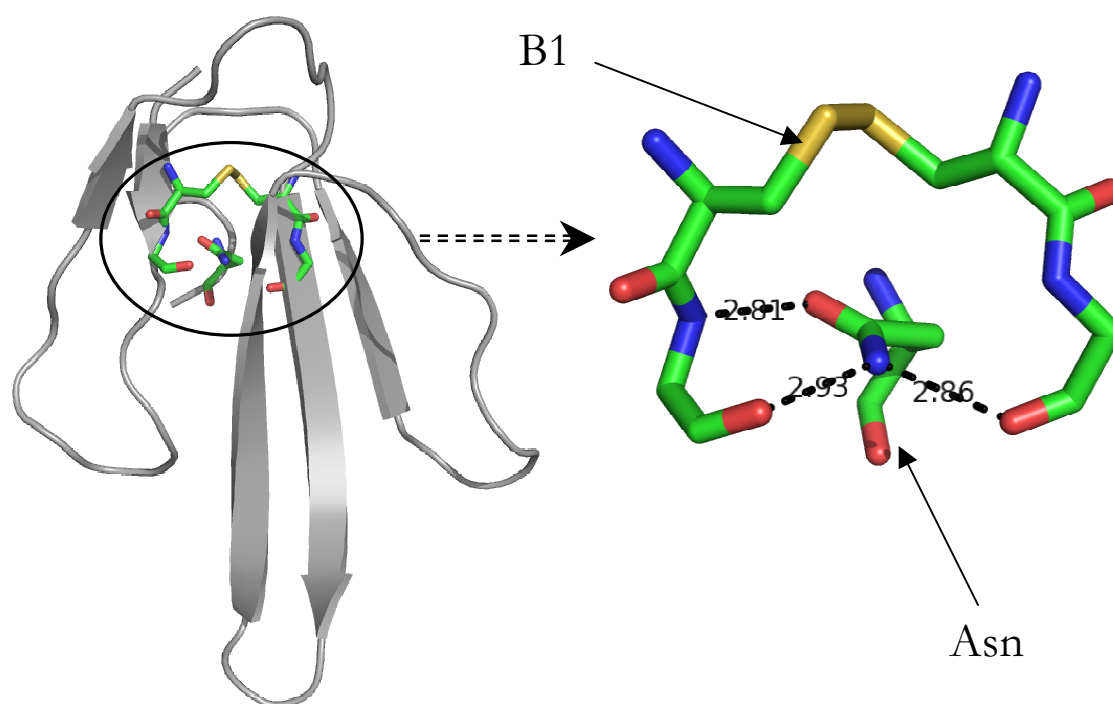


Figure 15 : Les 3 liaisons hydrogène impliquant la chaîne latérale de l'asparagine C-terminale conservée et le squelette peptidique de deux résidus adjacents aux cystéines du pont 1 (structure de l'atratoxine de *Naja atra* , code PDB : 1VB0).

En conclusion, les ponts 1, 2, 4, certains de leurs résidus adjacents et l'asparagine C-terminale sont conservés dans la séquence et forment un motif structural présent dans toutes les protéines de la famille trois doigts.

## 6 LE REPLIEMENT DES TOXINES A TROIS DOIGTS

### 6.1 Repliement des neurotoxines trois doigts

Les premières études concernant le repliement oxydant des toxines trois doigts datent du début des années 80 ((Ménez et al., 1980)). Ménez *et al.* ont étudié le repliement oxydant à l'air de différentes neurotoxines courtes naturelles dont la longueur de la séquence était de 60, 61 ou 62 résidus. Le suivi de la réaction de repliement était effectué par mesure de l'apparition des structures secondaires en dichroïsme circulaire. Les auteurs ont interprété les résultats grâce à un mécanisme à deux états et ont déterminé la vitesse initial de la réaction. Les valeurs de vitesse initiale de repliement sont présentés dans le Tableau 2.

Nom	Espèce	Nbre a.a.	Séquence	Vi
Toxine a	A. s.	60	MTCCNQSSQPKTTTNC-AG-NSCYKKTWSDHRGTIIERGCGCPQVKSGIKLECCHTNECNN	5,6
Toxine b	A. l.	60	LTCCNQSSQPKTTTDC-AD-NSCYKMTWRDHRGTIERGCGCPQVKPGIKLECCKTNECNN	5,6
Toxine $\alpha$	N. n.	61	LECHNQSSQPPTTKTC-PGETNCKYKKVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDCNN	4,0
Toxine d	N. m.	61	MECHNQSSQPPTTKTC-PGETNCKYKQWSDHRGTIIERGCGCPSVKKGKINCCTTDRCNN	2,9
Toxine IV	H. h.	61	LECHNQSSQPTTQTCT-PGETNCKYKQWSDHRGSRTERGCGCPTVKPGIKLNCCTTDRCK	6,4
Toxine $\alpha$	N. n. p.	61	LECHNQSSQAPTTKTC-SGETNCKYKQWSDHRGTIIERGCGCPKVKPGVKLNCCTTDRCNN	6,4
Erabutoxine a	L. s.	62	RICFNHQSSQPQTTKTCSPGESSCYNKQWSDFRGTIIERGCGCPTVKPGIKLSCCESEVCNN	0,8
Erabutoxine b	L. s.	62	RICFNHQSSQPQTTKTCSPGESSCYHKQWSDFRGTIIERGCGCPTVKPGIKLSCCESEVCNN	0,7
Cobrotoxine	N. n. a.	62	LECHNQSSQPTTTCGSGGETNCKYKRWDRHGYRTERGCGCPVKNGIEINCCTTDRCNN	0,6

Tableau 2 : Séquences et vitesses initiales de repliement de différentes neurotoxines courtes. V : vitesse initiale de repliement en  $10^{-5} \text{ s}^{-1}$ . L. s. : *Laticauda semifasciata*, N. n. a. : *Naja naja atra* ; N. n. : *Naja nigricollis* ; N. m. : *Naja melanoleuca* ; H. h. : *Hemachatus haemachatus* ; N. n. p. : *Naja naja philippinensis* ; A. s. : *Astrotia stokesii*, A. l. : *Aipysurus laevis*.

A partir de ces résultats, les auteurs ont proposé que les importantes différences de cinétique observées au sein de cette série de protéines étaient dues principalement à l'addition d'un acide aminé dans lk1 (voir Figure 11). Cependant l'existence d'autres différences dans ces séquences ne permettait pas d'affirmer que l'addition d'un résidu dans lk1 était l'unique facteur influençant la vitesse de repliement. Pour obtenir une réponse plus fiable à cette question, Ruoppolo *et al.* ((Ruoppolo et al., 1998)) ont étudié le repliement oxydant de variants de la toxine  $\alpha$  de *Naja nigricollis* en présence de glutathion oxydé et de glutathion réduit. L'apparition de la forme native a été suivie en analyse en spectrométrie de masse.

Nom	Séquence	1/2 vie
Toxine a	MTCCNQSSQPKTTTNC-AG-NSCYKKTWSDHRGTIIERGCGCPQVKSGIKLECCHTNECNN	<60
$\alpha 60$	LECHNQSSQPPTTKTC-GETNCKYKKVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDCNN	35
Toxine $\alpha$	LECHNQSSQPPTTKTC-PGETNCKYKKVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDCNN	60
Erabutoxine b	RICFNHQSSQPQTTKTCSPGESSCYHKQWSDFRGTIIERGCGCPTVKPGIKLSCCESEVCNN	210
$\alpha 62$	LECHNQSSQPPTTKTCSPGETNCKYKKVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDCNN	220

Tableau 3 : Vitesse de repliement de variants synthétiques de la toxine  $\alpha$  de *N. nigricollis*, et des toxines naturelles leur correspondant. D : Demi-vie de la forme native en minutes.

Cette fois-ci, les séquences des protéines comparées étaient identiques ( $\alpha 60$ ,  $\alpha 61$  et  $\alpha 62$ ), la seule différence étant la longueur de lk1. Les résultats obtenus montrent que le simple fait de d'ajouter un résidu dans lk1 induit une diminution de la vitesse de repliement d'un facteur 3,5 (Tableau 3). Une hypothèse mécanistique a été formulée par les auteurs pour expliquer l'influence de la longueur de lk1 sur la vitesse de repliement oxydant de ces molécules : une augmentation de la longueur de lk1 impliquerait une augmentation de la liberté conformationnelle de la chaîne dans cette région. Par conséquent, les possibilités d'appariements différents seraient accrues, et donc il se formerait un plus grand nombre d'intermédiaires de repliement dans les premières étapes. Le phénomène global de repliement serait alors ralenti.

## 6.2 Repliement de $\alpha 62$

A la suite de cette étude, la même équipe a détaillé le mécanisme de repliement de la toxine  $\alpha$  ((Ruoppolo et al., 2001)). Deux intermédiaires stables à trois ponts appelés C et D s'accumulent. L'attribution des ponts pour ces intermédiaires fut obtenue par digestion à la trypsine suivie d'une dégradation d'Edman couplée à l'analyse de masse des mélanges obtenus.

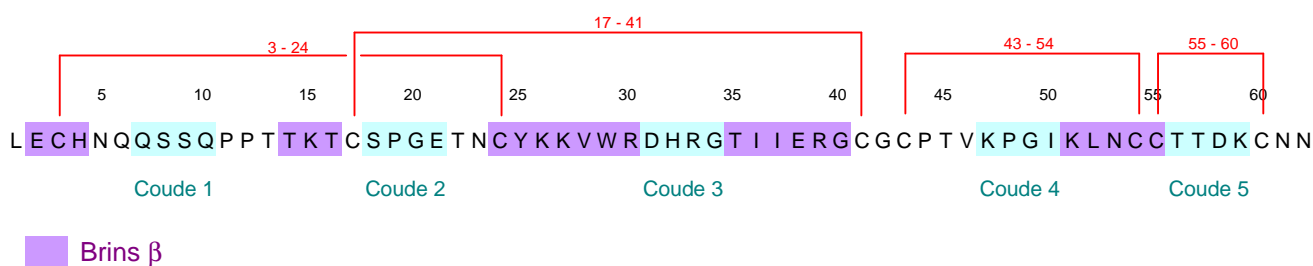


Figure 16 : Séquence, structure secondaire et ponts disulfure de  $\alpha 62$

Les travaux de Ruoppolo *et al.* ont permis de démontrer qu'il manque le pont [43-54] à l'intermédiaire C et le pont [17-41] à l'intermédiaire D (Figure 16). Les auteurs ont également bloqué à l'acide puis purifié ces intermédiaires en chromatographie phase inverse et dans un deuxième temps ont étudié le repliement oxydant à partir de ces intermédiaires purifiés. Ils ont pu alors proposer le mécanisme présenté sur la Figure 17.

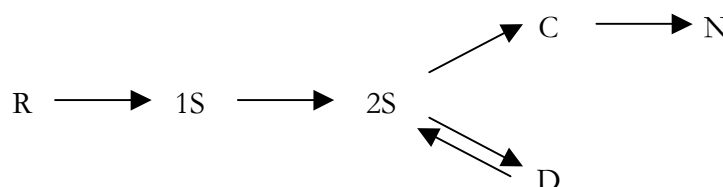


Figure 17 : Schéma du repliement oxydant de  $\alpha 62$  proposé par Ruoppolo *et al.*. Avec : R, forme dépliée réduite ; 1S, ensemble d'intermédiaires à un pont ; 2S, ensemble d'intermédiaires à deux ponts ; C, intermédiaire des[43,54] ; D, intermédiaire des[17,41] ; N, forme native à quatre ponts.

D'après ces résultats, lors du repliement de  $\alpha 62$ , l'intermédiaire D (ou des[17-41]) est une voie sans issue et il est nécessaire qu'un de ses ponts soit réduit pour former l'intermédiaire C qui lui, est productif. Les auteurs avaient remarqué que ce mécanisme est similaire à celui du BPTI, dans lequel des intermédiaires à ponts natifs constituent des pièges cinétiques. Comme pour le BPTI aucun intermédiaire « scrambled » (c'est à dire complètement oxydé mais dont l'appariement est non natif) n'a été mis en évidence.

### 6.2.1 Repliement des cardiotoxines

L'étude du repliement oxydant de la cardiotoxine III de *Naja naja atra* a montré que cette toxine se replie selon un mécanisme différent de celui de  $\alpha 62$  ((Chang et al., 1998), (Chang et al., 2006)). En effet, un massif d'intermédiaires possédant un nombre variable de ponts disulfure apparaît, aucun de ces intermédiaires n'étant majoritaire. De plus, lors de son repliement oxydant, cette protéine produit une quantité importante de formes à quatre ponts, donc complètement oxydées, mais dont l'activité est nulle. D'après le mécanisme proposé par Chang *et al.* ((Chang et al., 2006)), la conversion de ces intermédiaires à quatre ponts non natifs en la forme native constitue l'étape finale du repliement (Figure 18).

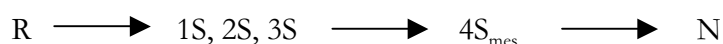


Figure 18 : Mécanisme de repliement de la cardiotoxine II proposé par Chang *et al.* ((Chang et al., 2006)). R : forme réduite ; 1S, 2S, 3S : formes à 1, 2 et 3 ponts ; 4S<sub>mes</sub> : forme à 4 ponts mesappariés ; N : forme native.

Comme cela a été rappelé en partie 5.2 de cette introduction, les mécanismes de repliement oxydant des petites protéines à ponts disulfure peuvent être classés en deux catégories ((Arolas et al., 2006)). Les résultats obtenus lors de l'étude de  $\alpha 62$  et de la cardiotoxine III montrent que la première appartient à la classe de type BPTI ((Arolas et al., 2006)) et la seconde à la classe de type hirudine ((Chang et al., 2006)), alors que ces deux protéines partagent des similitudes structurales. Dans le cas de la cardiotoxine III, il est possible que la présence de nombreux résidus hydrophobes nécessaires à la réalisation de la fonction conduise à un repliement de type « hydrophobic collapse ».



---

## 7 PROBLÉMATIQUES

---

L'étude du repliement de la toxine  $\alpha$  m'a conduit à m'intéresser à deux problématiques bien distinctes. Chacune de ces problématiques fait l'objet d'une partie dans les résultats et lors de la discussion.

### 7.1 Mécanisme de repliement *in vitro* de la toxine $\alpha$ de *N. nigricollis*

Les études qui ont décrit le repliement oxydant des neurotoxines courtes *in vitro* ont montré que la longueur de lk1 a un effet sur la cinétique de ces molécules. Les auteurs ont proposé que l'augmentation de la longueur de cette boucle influe sur la liberté conformationnelle dans les premières étapes et implique une augmentation du nombre d'intermédiaires formés. La première partie de ma thèse a pour but de tester cette hypothèse et de proposer une explication mécanistique et structurale aux modifications cinétiques observées entre les différents variants de la toxine  $\alpha$ .

### 7.2 Vectorisation *in vitro* du repliement de la toxine $\alpha$ de *N. nigricollis*

Certaines protéines nouvellement synthétisées dans les cellules des organismes vivants peuvent commencer à se replier alors qu'elles sont en cours d'élongation. Dans ce cas le repliement est vectorisé puisqu'il démarre dans la partie N-terminale pour finir dans la partie C-terminale. Il est donc légitime de se poser la question suivante : l'aspect vectoriel du repliement influe-t-il sur le son efficacité ? Certaines équipes ont commencé à l'étudier par simulation numérique (partie 2.3 de cette introduction), mais aucune tentative de réponse expérimentale à cette question n'est apparue dans la littérature à ma connaissance. Nous avons tenté de mettre au point un dispositif expérimental permettant de tester l'influence de la vectorisation sur le repliement des protéines. L'idée de base était de permettre à la molécule de se replier dans sa partie N-terminale indépendamment de sa partie C-terminale, celle-ci venant s'ajouter par la suite. La stratégie que nous avons mise en œuvre consiste, dans un premier temps à bloquer la protéine dans son état déplié par la liaison de plusieurs anticorps spécifiques des différentes parties de la séquence afin d'inhiber le repliement et de masquer les résidus au reste de la molécule puis dans un deuxième temps à lever cette inhibition depuis l'extrémité N-terminale vers l'extrémité C-terminale. (Figure 19 et commentaires)

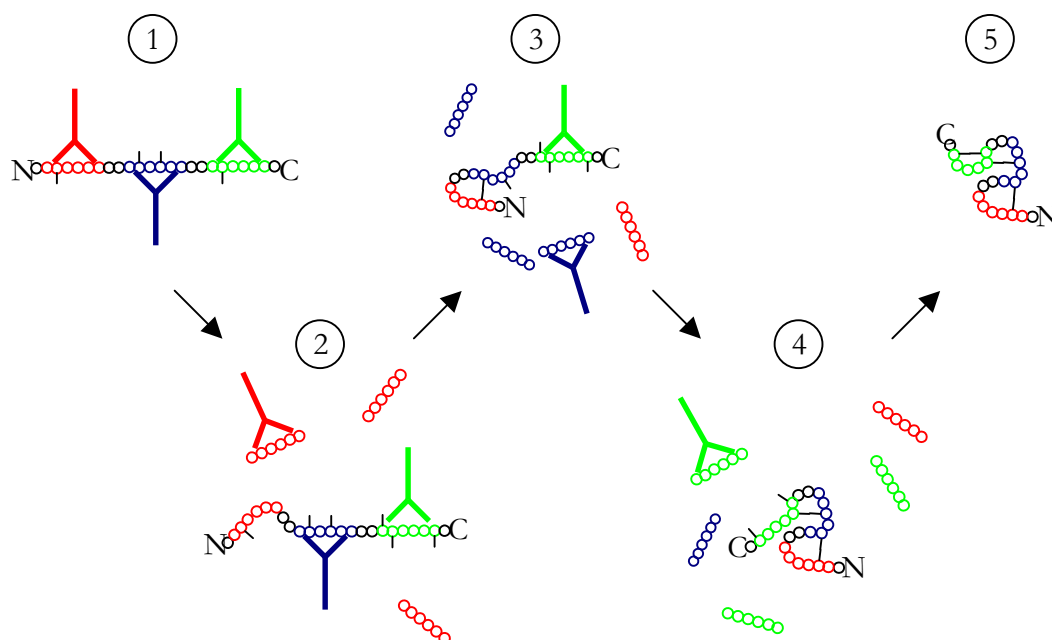


Figure 19 : schéma de principe de la vectorisation du repliement *in vitro*.

Etape 1 : La protéine réduite est dépliée et mise en présence de trois anticorps différents, chacun spécifique d'une partie de la séquence de la protéine étudiée. La présence de ces anticorps doit empêcher à la protéine de se replier même si le milieu contient le couple oxydo-réducteur permettant le repliement.

Etape 2 : Un large excès du peptide correspondant à l'épitope N terminal est ajouté. Par compétition, l'anticorps N terminal se décroche de la protéine entière à une vitesse correspondant à sa vitesse de dissociation. La partie N terminale n'est plus protégée par l'anticorps et peut alors commencer à se replier.

Etape 3 : Un large excès du peptide correspondant à l'épitope central est ajouté. De la même façon que pour l'anticorps N terminal, l'anticorps central se décroche de la protéine entière. Si deux cystéines impliquées dans un pont dans la forme finale sont libérées elles ont la possibilité de former ce pont alors que la partie C terminale reste dénaturée.

Etape 4 : Un large excès du peptide correspondant à l'épitope C terminal est ajouté. La partie C terminale est libérée. Les ponts présents dans cette partie peuvent se former.

Etape 5 : La protéine est repliée.

En utilisant la stratégie présentée ci-dessus pour vectoriser le repliement d'une protéine à ponts disulfure *in vitro*, le but est de permettre à des ponts natifs de se former dans la partie N-terminale de la molécule alors que les cystéines situées dans la partie C-terminale seraient masquées par les anticorps. Lorsque les protéines à ponts disulfure sont synthétisées *in vivo*, elles sont dans un deuxième temps transloquées dans le réticulum endoplasmique contenant un milieu oxydant qui va permettre l'oxydation des ponts. L'hypothèse testée dans cette étude est que la formation de ponts disulfure de façon cotranslocationnelle (qui implique une vectorisation de la formation des ponts disulfure) est plus efficace que lorsque les ponts disulfure ont la possibilité de se former simultanément dans n'importe quelle partie de la séquence, comme c'est le cas lors d'une expérience de repliement oxydant *in vitro* classique. Le but ultime est d'observer une augmentation de la vitesse lorsque le repliement est vectorisé.



# **MATERIELS ET METHODES**



# 1 PRODUCTION PAR VOIE RECOMBINANTE DES MUTANTS ETUDIÉS

## 1.1 Séquences et gènes des protéines étudiées

### 1.1.1 Séquences des protéines utilisées dans cette étude

Le tableau 4 présente les séquences des protéines étudiées. Tox61 est la toxine  $\alpha$  naturelle de *Naja nigricollis*. Tox60 et Tox62 sont des mutants de la toxine  $\alpha$  dans lesquels un résidu a été supprimé ou ajouté dans le linker 1 (voir figure 11 en partie 5.3.1 de l'introduction) et correspondent aux mutants  $\alpha 60$  et  $\alpha 62$  étudiés par Ruoppolo et al. ((Ruoppolo et al., 1998)). Tox61pT3 et Tox62pT3 sont des analogues structuraux des intermédiaires C de Tox61 et de Tox62 dans lesquels les cystéines libres dans ces intermédiaires (en l'occurrence celles du 3<sup>ème</sup> pont dans la séquence) ont été substituées par des alanines.

Dans la deuxième partie de cette étude, qui concerne la vectorisation du mécanisme de repliement, la protéine Tox62 a été utilisée comme modèle. Tox62 polyS est un mime de Tox62 sous sa forme réduite dans laquelle toutes les cystéines ont été mutées en sérine. Tox62 polySC a été utilisée pour ajouter une molécule de biotine à l'extrémité C terminale afin de l'immobiliser dans le cadre des expériences d'ELISA ou de Biacore en liant la biotine à une molécule de streptavidine elle-même immobilisée. Enfin, face à la difficulté d'obtenir des anticorps spécifiques de la partie C-terminale de Tox62, une séquence Etag, connue pour être fortement immunogène a été ajoutée à l'extrémité C-terminale (Tox62 Etag). Les mutants Tox62 polySETag et Tox62 polySETagC ont été produits pour les mêmes raisons que Tox62 polyS et Tox62 polySC.

Nom	Séquence
Tox60	LECHNQSSQPPTTKTC-GETNCKYKVVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDKCNN
Tox61	LECHNQSSQPPTTKTC-PGETNCKYKVVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDKCNN
Tox62	LECHNQSSQPPTTKTC-SPGETNCKYKVVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDKCNN
Tox61pT3	LECHNQSSQPPTTKTC-PGETNCKYKVVWRDHRGTIIERGCGAPTVPKPGIKLNCCTTDKCNN
Tox62pT3	LECHNQSSQPPTTKTC-SPGETNCKYKVVWRDHRGTIIERGCGAPTVPKPGIKLNCCTTDKCNN
Tox62 polyS	LESHNQSSQPPTTKTSSPGETNSYKKVWRDHRGTIIERGSSTPTVPKPGIKLNCCTTDKSNNC
Tox62 polySC	LESHNQSSQPPTTKTSSPGETNSYKKVWRDHRGTIIERGSSTPTVPKPGIKLNCCTTDKSNNC
Tox62 Etag	LECHNQSSQPPTTKTCCPGETNCKYKVVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDKCNNGAPVPYPD
Tox62 polySETag	LESHNQSSQPPTTKTSSPGETNSYKKVWRDHRGTIIERGSSTPTVPKPGIKLNCCTTDKSNNGAPVPYPD
Tox62 polySETagC	CLESHNQSSQPPTTKTSSPGETNSYKKVWRDHRGTIIERGSSTPTVPKPGIKLNCCTTDKSNNGAPVPYPD

Tableau 4 : Noms et séquences des protéines utilisées dans cette étude. Les résidus modifiés par rapport à la toxine  $\alpha$  naturelle (appelée ici Tox61) sont grisés.

Les noms des protéines Tox62 polyS, Tox62 polySC, Tox62 polySETag et Tox62 polySETagC seront simplifiés dans le reste de l'étude en polyS, polySC, polySETag et polySETagC respectivement.

Toutes ces protéines ont été produites par voie recombinante chez *Escherichia coli* comme décrit ci-après. Seule la toxine Tox60 a été obtenue par synthèse chimique : la stratégie Fmoc/tBu a été mise en œuvre sur un appareil de synthèse Applied Biosystems 433 selon le protocole établi par le constructeur.



Ce vecteur permet l'insertion d'un gène entre les sites de restriction KpnI et BamHI. La séquence correspondant à la toxine est ainsi fusionnée à celle de la protéine ZZ permettant la production d'une toxine hybride ZZ-Tx sous la dépendance d'un promoteur transcrit uniquement par l'ARN polymérase du phage T7(P<sub>10</sub>ØT7). Il convient de noter qu'un résidu méthionine occupe la position -1 par rapport au résidu N-terminal de la toxine.

### 1.3.2 Préparation du vecteur

Le vecteur pCP est digéré successivement par deux enzymes de restriction BamHI et KpnI. Il est ensuite déphosphorylé par l'utilisation de l'enzyme CIP (Calf Intestine Phosphatase).

### 1.3.3 Préparation de l'insertion

L'insertion est digérée successivement par les deux enzymes de restriction BamHI et KpnI.

### 1.3.4 Réaction de ligation

La réaction de ligation est effectuée par la mise en présence du vecteur (50 ng), de l'insert (en excès molaire 5 ou 10x soit 12 ou 24 ng), de la ligase et de son tampon. La réaction se fait dans un volume de 50 µL à température ambiante pendant 10 minutes.

### 1.3.5 Transformation des bactéries

Immédiatement après la ligation, des bactéries E. Coli XL1 Blue sont transformées. 2 µL de réaction de ligation sont dilués dans 50 µL de bactéries compétentes. Les bactéries sont incubées 30 minutes dans la glace, subissent un choc thermique à 42°C, puis sont remises à incuber 2 minutes dans la glace. La culture est diluée dans 950 µL de milieu SOC puis incubée 1 heure à 37°C. 100 µL de culture sont ensuite étalés sur une boîte de LB contenant de l'ampicilline à 200mg/mL. Les boîtes sont incubées une nuit à 37°C.

### 1.3.6 Vérification des clones

Les clones obtenus sont utilisés pour réaliser une PCR sur colonie permettant d'amplifier le gène qui a été intégré. On prépare dans ce cas un mélange de PCR identique à celui préparé pour la mutagenèse à ceci près que l'enzyme utilisée est une thermopolymérase (Biolabs) et le mélange réactionnel ne contient pas de matrice. Une fois le mélange préparé, les bactéries sont prélevées puis inoculées dans ce mélange. La matrice est contenue dans ces bactéries dans le cas où elles ont intégré le plasmide et le gène. Le programme utilisé est le même que pour la mutagenèse avec une étape supplémentaire (5min à 95°C) destinée à casser les cellules et à libérer leur contenu en acides nucléiques.

Après vérification par électrophorèse sur gel d'agarose 2% de la taille des fragments d'ADN amplifiés, il est possible de déterminer les clones ayant intégré un vecteur contenant le gène de la toxine.

### 1.3.7 Miniprep sur la colonie vérifiée

Les clones sélectionnés sont mis en culture dans 10mL de milieu LB liquide pour la nuit. L'utilisation d'un kit d'extraction (miniprep Wizard plus Promega) permet d'extraire l'ADN de ces cultures. Le plasmide purifié est précipité à l'éthanol, séché, et repris dans de l'eau PPI (Pour Préparation Injectables) puis stocké à -20°C.

### 1.3.8 Vérification de la séquence

La séquence du gène intégré dans ces clones est vérifiée par l'utilisation de la méthode des « Big Dye Terminator » (Applied Biosystem). Les plasmides correspondant aux clones vérifiés sont conservés à -20°C.



## 1.4 Expression et purification

### 1.4.1 Stratégie de production et de purification

Drevet *et al.* ((Drevet et al., 1997)) avaient démontré la possibilité de produire une neurotoxine de serpent (Erabutoxine a) chez *E. coli* en utilisant le système d'expression DE3 LysS associé au vecteur pCP. Nous avons utilisé ce système d'expression dans cette étude. La Figure 21 présente les différents éléments contenus dans le cytoplasme d'une bactérie BL21(DE3) Lys S.

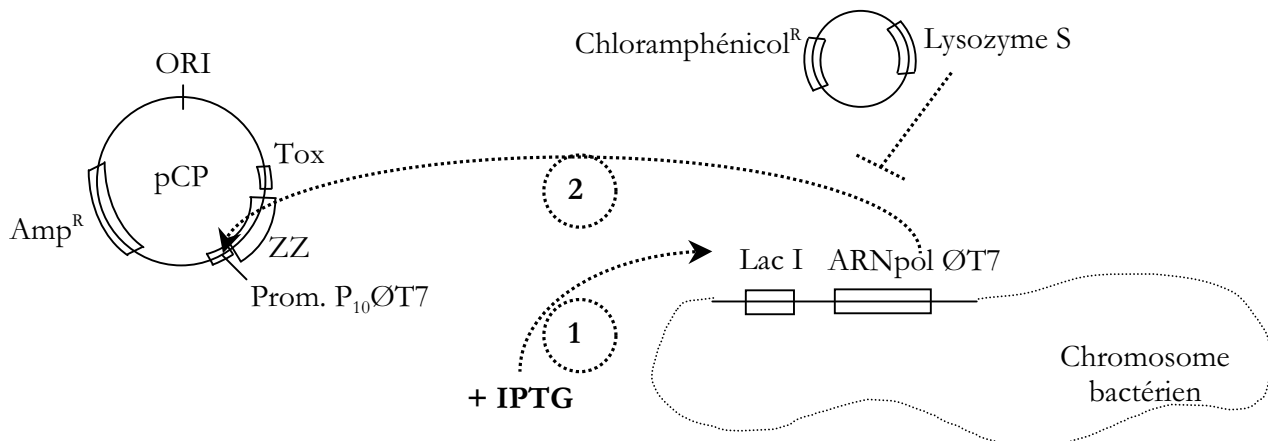


Figure 21 : Système d'expression DE3 LysS associé au vecteur pCP.

Les bactéries DE3 ont été conçues pour exprimer un gène cloné, sélectivement et avec un haut niveau d'expression. Le gène de l'ARN polymérase du phage T7 a été incorporé dans le chromosome de ces bactéries sous contrôle de LacI ((Studier and Moffatt, 1986)). Cette polymérase a la particularité d'être hautement sélective des promoteurs du phage T7. Notre gène est placé sous contrôle du promoteur de la protéine 10 du phage T7 (P<sub>10</sub>ØT7). Pour diminuer la fuite d'expression liée à ce système, un deuxième plasmide contenant le lysozyme du phage T7 a été ajouté ((Studier, 1991)). Ce lysozyme a la propriété d'inhiber l'action de l'ARN polymérase du phage T7 (DE3 LysS). L'ajout d'IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside) dans le milieu de culture induit l'activation du gène de l'ARN polymérase du phage T7. Le promoteur P<sub>10</sub>ØT7 contenu dans le vecteur pCP permet l'initiation de la transcription du gène de la toxine par l'ARNpolymérase du phage T7.

Le plasmide pCP contient la fusion de la toxine à la protéine ZZ. Cette protéine est composée de deux domaines synthétiques Z, dérivés du domaine B de la protéine A du staphylocoque ((Nilsson et al., 1987)). Les avantages de ce système sont que d'une part la protéine ZZ est très bien exprimée chez *E. coli*. Une fois libérée dans le cytoplasme bactérien, la fusion est moins sujette à la dégradation et est plus soluble que la toxine seule. D'autre part la protéine ZZ ayant la propriété de se lier aux IgG (Immunoglobulines de classe G), elle fournit une méthode simple et efficace de purification de la fusion par chromatographie d'affinité sur colonne d'IgG-Sepharose. Enfin comme il n'existe qu'une méthionine en -1 de la toxine, il est possible de séparer la toxine de la protéine de fusion par coupure au bromure de cyanogène.

## 1.4.2 Protocole de production et de purification

### 1.4.2.1 Transformation et culture

Les bactéries DE3 Lys S sont transformées avec le plasmide pCP contenant le gène de la protéine que l'on souhaite produire. A 50  $\mu\text{L}$  de bactéries compétentes on ajoute 20 ng de vecteur. Le mélange est incubé 30 minutes dans la glace puis subit un choc thermique de 1 minute à 42°C et est remis à incuber dans la glace. Le mélange est ensuite dilué 20 fois dans du milieu SOC puis incubé une heure à 37°C. 100 $\mu\text{L}$  de culture sont prélevés puis étalés sur une boîte LB Agar contenant de l'ampicilline (200 $\mu\text{g}/\text{mL}$ ) et du chloramphénicol (30 $\mu\text{g}/\text{mL}$ ). Les boîtes sont incubées une nuit à 37°C. Lors de la pré-culture, une colonie isolée est mise en culture dans 10 mL de milieu LB contenant de l'ampicilline (200 $\mu\text{g}/\text{mL}$ ) et du chloramphénicol (30 $\mu\text{g}/\text{mL}$ ). Le tube est incubé une nuit à 37°C. La culture d'expression est effectuée à partir de 2 mL de pré-culture dilués dans 200mL de milieu LB contenant de l'ampicilline (200 $\mu\text{g}/\text{mL}$ ) et du chloramphénicol (30 $\mu\text{g}/\text{mL}$ ). Elle est effectuée à 37°C pendant environ 2h30. Lorsque la turbidité mesurée à 600nm atteint environ 0,8, l'IPTG est ajouté afin d'atteindre une concentration finale de 1mM. Le mélange est incubé pendant 4 heures à 37°C. Les cultures sont centrifugées pendant 30 minutes à 10 000 tr/min à 4°C. Le culot est repris dans du tampon PBS 1X avec EDTA 1mM et DTT 10mM. Les suspensions cellulaires sont congelées pour la nuit.

### 1.4.2.2 Extraction

Une fois les suspensions décongelées, ces-dernières sont recongelées dans un mélange d'éthanol et de carboglace pendant 20 minutes puis passées à la presse Eaton. Le mélange est soumis aux ultrasons pour casser les acides nucléiques. Du PMSF et du sulfate de protamine sont ajoutés aux cellules cassées pour atteindre une concentration de respectivement 0,5mM et 2 mg/L. Le mélange est incubé 10 à 15 minutes dans la glace puis centrifugé 10 minutes à 4°C à 10 000 tr/min. Le surnageant est récolté.

### 1.4.2.3 Chromatographie d'affinité et repliement

L'extrait protéique est incubé sous agitation en présence d'IgG-sepharose dans du PBS contenant de l'EDTA à 1mM pendant 30 minutes à 4°C. Le mélange est coulé dans une colonne de verre (Biorad). La matrice est ensuite lavée avec du tampon PBS +EDTA 1mM jusqu'à élimination de la fraction non retenue. La colonne est ensuite rincée avec 2 volumes de tampon de repliement (Tris-HCl 100mM pH 7,6, EDTA 1mM, GSH/GSSG : 3/0,3mM) puis incubée 48 heures à 20°C. La colonne est ensuite lavée en PBS, puis les protéines retenues sont éluées en acide acétique pH 3,4. La fraction éluee contenant la fusion que l'on appellera ZZtox est purifiée sur HPLC phase inverse sur colonne C5 (Supelco, 250x10mm). Le tampon A est une solution d'acide trifluoroacétique (TFA) 0,1%vol dans l'eau. Le tampon B contient de l'acétonitrile à 90%vol dans l'eau et du TFA à 0,1%vol. Le gradient permet de passer de 18 à 35 % d'acétonitrile en 30 minutes à un débit de 3 mL/min. Le pic correspondant à la fusion est récolté puis lyophilisé.

#### 1.4.2.4 Coupure au bromure de cyanogène et purification de la toxine

La séparation de la protéine de fusion ZZ de la toxine est effectuée par coupure au bromure de cyanogène (CNBr). Le mélange réactionnel contient : 100 excès molaires de CNBr, 100mM de tryptophane et de l'acide chlorhydrique à 0,1 N. La présence de tryptophane dans le milieu permet d'éviter une dégradation excessive des tryptophanes des toxines sous l'action de l'acide. Le mélange est ensuite incubé 6 heures à température ambiante et à l'abri de la lumière puis dessalé sur une colonne de gel filtration PD10 (GE Healthcare). La purification de l'échantillon est effectuée en HPLC phase inverse sur colonne C5 (Supelco, 250x10mm). On utilise alors un protocole identique à celui de la purification de ZZtox. La fraction susceptible de contenir la toxine est récoltée puis lyophilisée. Un échantillon de la protéine isolée est ensuite analysé par spectrométrie de masse afin de vérifier que la protéine produite a une masse identique à celle attendue.

#### 1.4.2.5 Rendements de production et de purification

Pour la production de la protéine Tox62, à partir d'un litre de culture, on obtient 83 mg de fusion ZZ/toxine, ce qui correspond à 28 mg de toxine. Après coupure et purification en HPLC, 8,5 mg de Tox62 pure sont récupérés.

Les rendements de production obtenus pour les autres mutants sont comparables à ceux de Tox62. Pour les mutants de ponts, des étapes supplémentaires de purification ont été nécessaires, les rendements de purification de ces mutants étaient donc fortement inférieurs (3 mg de toxine pure pour 1L)

---

## 2 MARQUAGE DES TOXINES

---

### 2.1 Marquage à la biotine

Une solution contenant le mutant polySC (voir Tableau 4) à une concentration 1,5  $\mu\text{M}$  dans un tampon 50 mM Tris-HCl pH 7,8 est préparée. De la biotine-BMCC (Pierce) est ajoutée pour atteindre une concentration de 0,4 mM. Après deux heures d'incubation à température ambiante, Tx62polySC-biot est purifié par HPLC phase inverse sur colonne C5(Supelco, 250x10mm) en utilisant le gradient utilisé pour purifier ZZTox (voir ci-dessus). La molécule obtenue est appelée polySC-biot. Le même protocole est utilisé pour produire polySETAGC-biot à partir de polySETAGC.

### 2.2 Marquage à l'azote 15

Les protéines Tox61, Tox62, Tox61pT3 et Tox62pT3 utilisées lors de l'étude en RMN ont été marquées à l'azote 15. Pour produire ces protéines marquées  $^{15}\text{N}$ , le protocole décrit en partie 1 de ces matériels et méthodes est utilisé ; lors de la préculture et de la culture le milieu LB est remplacé par un milieu minimum M9 préparé avec du sulfate d'ammonium  $^{15}\text{N}$  (Isotec, Matheson USA ).

### 2.3 Marquage au tritium

La tyrosine de la protéine Tox62 a été tritiée par Marie-Anne Lelait (SIMOPRO CEA-Saclay). La première étape consiste à réaliser une iodation de la tyrosine de la protéine par ajout d'ICl en excès dans un tampon Tris-HCl 0,5M pH8. La protéine est dessalée par gel filtration sur résine P2 (Bio-Gel) dans l'acide acétique 10% dans l'eau. La solution obtenue est lyophilisée. La protéine iodée lyophilisée est reprise dans un tampon phosphate 0,1M pH 7,2. La tritiation est alors effectuée en milieu hétérogène en présence du catalyseur  $\text{PdAl}_2\text{O}_3$  et d'une surpression de  $^3\text{H}_2$ . La réaction est menée pendant 30 minutes sous agitation à température ambiante. Le  $\text{PdAl}_2\text{O}_3$  est éliminé par centrifugation. Le surnageant est évaporé au Speedvac. Des cycles de remise en solution suivie d'une évaporation sont effectués pour éliminer le  $^3\text{H}_2$  non lié contenu dans le tampon. La protéine est ensuite purifiée par HPLC phase inverse sur une colonne C18 Atlantis (Waters). Le tampon A utilisé contient 0,1% d'acide formique dans l'eau et le tampon B 0,1% d'acide formique dans l'acétonitrile. Le gradient permet de passer de 21 à 27% d'acétonitrile en 19 minutes. La fraction correspondant à la toxine est évaporée au Speedvac puis reprise dans du tampon PBS. Cette solution a une concentration en toxine de 128  $\mu\text{M}$  et une activité spécifique de 14 700Ci/mol.

### 2.4 Marquage au soufre 35

Des cellules *Escherichia coli* BL21(DE3) sont transformées avec le plasmide pCPTx62ETag (voir partie 1.1 de ces matériels et méthodes) puis mises en culture pour la nuit à 37°C dans un milieu minimum M9 contenant de l'ampicilline et du chloramphénicol. 500  $\mu\text{l}$  de cette préculture sont utilisés pour semer 30mL de milieu minimal M9 contenant les antibiotiques. La croissance est suivie jusqu'à l'obtention d'une turbidité à 600nm de 0,6. De l'IPTG est alors ajouté pour atteindre une concentration de 1mM et la culture est incubée 1 heure. 5mCi de cystéine 35S (Perkin Elmer) sont alors ajoutés. La culture est incubée encore pendant une heure avant que l'extraction et la purification de la toxine ne soient réalisées comme décrit en partie 1.4 de ces matériels et méthodes. Le jour de la production, la protéine a une activité spécifique de 4 200 Ci/mol et une concentration en protéine totale de 3 $\mu\text{M}$ .

### 3 REPLIEMENT OXYDANT *IN VITRO* DE LA TOXINE A

#### 3.1 Réaction de repliement

La protéine oxydée, purifiée et lyophilisée (1mg pour les temps courts, 200µg pour les temps longs) est reprise dans une solution de dénaturation (excès de DTT 50 fois mol/mol de cystéines et 6 M de chlorure de guanidium) puis incubée pendant deux heures à 37°C pour dénaturation. La protéine réduite et dénaturée est dessalée en condition acide par HPLC phase inverse sur colonne C18 (Phenomenex Jupiter 150x4,6mm). Le tampon A est une solution d'acide trifluoroacétique (TFA) 0,1%vol dans l'eau. Le tampon B contient de l'acétonitrile à 60%vol dans l'eau et du TFA à 0,1%vol. Le gradient permet de passer de 0 à 54 % d'acétonitrile en 18 minutes à un débit de 1 mL/min. Le pic correspondant à la protéine réduite est récolté puis lyophilisé pendant la nuit. Afin d'éviter une renaturation partielle de la protéine, le lyophilisat est repris dans de l'acide acétique 0,01% et utilisé immédiatement. En effet, comme cela a été décrit en partie 3.1.1 de l'introduction, abaisser le pH permet de prévenir la formation d'ion thiolate et donc toute oxydation des ponts disulfure. Les produits utilisés pour cette réaction sont saturés d'argon afin d'éliminer le dioxygène de l'air dissout dans ces solutions.

Le repliement est effectué dans une solution composée de tampon Tris-HCl à 0,1 M , d'EDTA à 1mM, pH 7,5, du glutathion réduit (GSH) à 3mM, du glutathion oxydé (GSSG) à 0,3 mM. La concentration en toxine réduite est de 17 µM. Le pH du mélange réactionnel est vérifié et éventuellement corrigé en ajoutant quelques microlitres d'une solution de Tris à 2M . La réaction est effectuée à 25°C. A différents temps, une fraction aliquote est prélevée. La réaction est bloquée soit par abaissement du pH à l'aide d'acide (solution d'acide trifluoroacétique à 0,1%vol), soit par ajout d'un agent bloquant les thiols libres. L'échantillon obtenu est plongé dans l'azote liquide puis conservé à -20°C avant analyse.

#### 3.2 Analyse des mélanges réactionnels

Le type d'information que l'on souhaite obtenir définit le type d'expérience effectué. Cependant, ce qui a été décrit ci-dessus est commun à toutes les expériences de repliement oxydant.

##### *3.2.1 Suivi des étapes initiales du repliement en chromatographie d'échange d'ions*

Dans ce cas l'objectif est d'obtenir des informations sur les premières étapes du repliement. Lors de ces étapes le nombre d'espèces existant dans le mélange de repliement est grand, il est donc nécessaire de réaliser le repliement d'une quantité importante de protéine réduite (en l'occurrence 1 mg), pour obtenir une quantité détectable de chacune des espèces. Toutes les informations données en partie 3.1 ci-dessus sont valables et le volume réactionnel total est multiplié par 10 (soit 6mL) afin de garder la même concentration en toxine tout en multipliant par 10 la quantité contenue dans chaque prélèvement, à savoir 10 nmol.

On s'intéresse aux étapes initiales du repliement oxydant. Des échantillons sont prélevés après 0, 1, 2, 4, 6, 8 , 10, 15, 20, 30 minutes. Dans le but de réaliser un prélèvement à 0 minute le plus précoce possible, un dispositif spécial a été mis au point. Le repliement est effectué dans un bécher sous agitation (nécessairement un petit peu de dioxygène est incorporé à cette étape, l'agitation est donc activée extemporanément). Au temps 0, la toxine réduite est ajoutée avec une micropipette et simultanément avec une autre micropipette le premier prélèvement est effectué. Le temps 0 minutes correspond à une à deux secondes de repliement.

L'objectif est de séparer les espèces à faible nombre de ponts qui apparaissent dans les premières étapes du repliement. Pour séparer les espèces on utilise une stratégie consistant à bloquer les cystéines libres avec un agent bloquant possédant une charge. Ainsi, en fonction du nombre de cystéines libres et donc en fonction du nombre de charges incorporées, le comportement chromatographique en échange d'ion est modifié. Ici on utilise le MTSBS (Sulfonato Butyl Méthane Thio Sulfonate fournit par la société TRC : Toronto Research Chemicals) qui permet l'incorporation d'une charge négative et qui a été décrit comme très réactif ((Kenyon and Bruice, 1977)). Pour le blocage, à chaque prélèvement de 580µL, 11,6µL de MTSBS 1M dans du DMSO (afin de prévenir son hydrolyse) sont ajoutés. Le mélange est incubé 5 minutes à température ambiante puis plongé dans l'azote liquide et conservé à -20°C jusqu'à son analyse.

Avant analyse en HPLC, les échantillons sont décongelés. Le tampon de réaction est remplacé par le tampon A d'échange d'ion (détaillé ci-après) grâce à une colonne de gel filtration PD10. L'échantillon obtenu est injecté sur une colonne d'échange de cations Mitsubishi MCI Gel, ProtEX SP 4,6x50mm. Le tampon A est un tampon citrate à 20mM pH 3, le tampon B est identique au tampon A mais contient en plus 2M de NaCl. L'ensemble de la chromatographie se déroule à un débit de 0,6 mL/min. Le gradient d'analyse est le suivant : après injection, la colonne est lavée pendant 20 minutes. On passe ensuite de 0 à 0,6 M de NaCl en 30 minutes. Des fractions de 200µL sont prélevées automatiquement tout au long du gradient. Ces fractions sont regroupées par pic d'élution, plongées dans l'azote liquide et conservées à -20°C.

### 3.2.2 Suivi des étapes finales du repliement en HPLC phase inverse

Lors des étapes finales du repliement oxydant de la toxine, les intermédiaires qui s'accumulent sont moins nombreux ce qui implique que des prélèvements de 1nmol de toxine sont suffisants. La réaction de repliement est alors stoppée, par ajout de 9 volumes d'acide trifluoroacétique 0,1% volume dans l'eau. L'échantillon est ensuite immédiatement plongé dans l'azote liquide.

Les différents mélanges réactionnels stabilisés sont décongelés dans les jours suivants la réaction de repliement pour analyse en HPLC phase inverse sur colonne C18 (Merck Chromolith 50 x 4,6mm). Le tampon A est une solution d'acide trifluoroacétique (TFA) 0,1%vol dans l'eau. Le tampon B contient de l'acétonitrile à 60%vol dans l'eau et du TFA à 0,1%vol. Pour le cas de Tox62, le gradient d'analyse est le suivant : on passe de 12 à 36% d'acétonitrile en 10 minutes à un débit de 3 mL/min.

Dans le cas du repliement oxydant de Tox60 et Tox61, le gradient commence par un palier de 6 minutes à 18% d'acétonitrile puis on passe de 18% à 27% d'acétonitrile en 13 minutes à un débit de 1mL/min.

## 3.3 Modélisation des cinétiques

Le logiciel Madonna Berkeley<sup>TM</sup> (<http://www.berkeleymadonna.com/>) a permis la résolution numérique des systèmes d'équations différentielles décrivant l'évolution des différentes espèces et d'optimiser les valeurs des constantes cinétiques pour s'approcher des résultats expérimentaux.

### 3.4 Analyse des formes intermédiaires

#### 3.4.1 Purification des intermédiaires C et D de Tox61 et de Tox62

Afin de produire une quantité importante d'intermédiaires C et D, 2mg de Tox61 et de Tox62 sont dénaturés selon le protocole décrit en partie 3.1 de ces matériels et méthodes. Les toxines Tox61 et Tox62 réduites obtenues sont repliées dans les conditions décrites en partie 3.1 de ces matériels et méthodes. Les repliements de Tox61 et de Tox62 après respectivement 30 minutes et 1 heure sont stoppés par ajout de 9 volumes d'acide trifluoroacétique 0,1% volume dans l'eau. L'échantillon est ensuite immédiatement plongé dans l'azote liquide. Les fractions C et D de chacune des protéines sont purifiées sur colonne d'HPLC phase inverse (Phenomenex Jupiter C18). Le débit utilisé est de 1mL/min. Les tampons utilisés sont ceux décrits en partie 3.1 de ces matériels et méthodes. Le gradient permet de passer de 19% à 21% d'acétonitrile en 16 minutes à un débit de 1 mL/min. Les pics C et D de chacune des toxines sont lyophilisées sur la nuit.

#### 3.4.2 Repliement des intermédiaires C et D de Tox61 et de Tox62

Les intermédiaires C et D purifiées comme décrit ci-dessus sont utilisés comme molécule de départ pour une expérience de repliement. Le même protocole que celui décrit en partie 3.1 de ces matériels et méthodes est utilisé. Le but de ces expériences étant de déterminer si ces intermédiaires sont productifs, le tampon de repliement utilisé ne contient pas de glutathion réduit. Le reste du protocole est parfaitement identique à celui présenté en partie 3.1 de ces matériels et méthodes. Les prélèvements sont effectués pour Tox61 C et D après 0', 2', 5', 10', 15', 30', 1h, 2h, 4h, 6h, 24h ; pour Tox62 C et D après 0', 30', 1h, 2h, 4h, 6h, 24h de repliement. Les échantillons obtenus sont analysés en HPLC RP comme décrit en partie 3.1 de ces matériels et méthodes avec le gradient 12 à 36%.

#### 3.4.3 Blocage au iodoacétamide

Après dénaturation, repliement, blocage à l'acide, purification et lyophilisation des intermédiaires C et D de Tox61 et Tox62, comme décrit ci-dessus, ces intermédiaires sont repris dans 200µL de solution de blocage (iodoacétamide 1M dans du tampon Tris-HCl 100mM pH8) puis incubées 1 minute à température ambiante. Une colonne de gel filtration PD10 (GE Healthcare) permet de remplacer le tampon de la solution de blocage par de l'eau. La solution obtenue est lyophilisée pendant la nuit.

#### 3.4.4 Identification des intermédiaires par analyse de masse

Toutes les analyses de masse décrites ci-après ont été réalisées à l'aide d'un spectromètre de masse MALDI TOF/TOF 4800 (Applied Biosystems), par l'équipe de Jean Pierre Le Caer de l'Institut de Chimie des Substances Naturelles (ICSN, CNRS Gif-sur-Yvette).

Lors des repliements de Tox61 et de Tox62, deux intermédiaires s'accumulent dans les étapes finales du repliement. Ruoppolo *et al.* ont caractérisé de manière approfondie les intermédiaires de Tox62 ((Ruoppolo et al., 2001)). Dans le cadre de cette étude, pour le cas de Tox61, les intermédiaires apparaissant dans les étapes finales du repliement ont été purifiés, bloqués, dessalés et lyophilisés comme décrit précédemment puis leur masse a été obtenue par MALDI-TOF, en mode linéaire positif avec comme matrice l'acide sinapinique.

Pour déterminer les cystéines libres dans ces intermédiaires, une digestion trypsique a été effectuée dans les conditions suivantes. La quantité d'intermédiaire a été ajustée pour chaque échantillon à 2µg. La digestion est réalisée dans 100 µl d'acétate d'ammonium 0,1M pH 6,5. 0,2µg de trypsine ont été ajoutés et la digestion est menée pendant 2 heures à 37°C dans un incubateur à sec (Thermomixer). Les mélanges de digestion obtenus sont réduits pendant 2 heures à 37°C par ajout d'une concentration de 2,5 mM de TCEP soit un excès 100 fois par rapport au nombre de cystéines des toxines. Les mélanges obtenus ont été analysés en MALDI-TOF en mode

réflectron positif avec une matrice 4 HCCA ( 4 hydroxy alpha sinamic acid). Les séquençages de certains des peptides ont été obtenus en spectrométrie de masse en mode tandem avec des énergies de collision de 1 ou 2 KeV avec l'air comme gaz de collision.

### 3.4.5 Dichroïsme circulaire

Les expériences de dichroïsme circulaire ont été effectuées sur un spectrophotomètre Jasco J-815. Toutes les expériences ont été menées avec une concentration en protéine de 20  $\mu$ M diluée dans un tampon phosphate à 0,4 mM pH 7,5.

#### 3.4.5.1 Acquisition de spectres de dichroïsme circulaire :

Les spectres ont été acquis à une température de 20°C, entre 190 et 250 nm avec une mesure tous les 0,5 nm. 5 spectres ont été accumulés. Avant chaque mesure un spectre est effectué sur du tampon, ce spectre est ensuite soustrait au spectre de la protéine.

#### 3.4.5.2 Dénaturation thermique suivie par mesure de l'ellipticité

Les toxines sont des protéines en feuillet  $\beta$ . Le spectre de dichroïsme circulaire d'une protéine en feuillet  $\beta$  contient une bande positive intense à 198 nm. Afin de suivre l'évolution de la structure secondaire lors de la dénaturation thermique, l'ellipticité à 198 nm est enregistrée à différentes températures. Une mesure tous les 0,2°C est acquise entre 4 et 70°C avec un gradient de température de 5°C par minute.

### 3.4.6 Cristallographie de Tox62

L'ensemble de ces expériences a été effectué par Enrico Stura et Renée Menez (SIMOPRO CEA-Saclay). Les cristaux ont été obtenus par la méthode de diffusion de vapeur en goutte assise. Les conditions de cristallisation sont : tampon citrate de sodium 100mM pH 4,5, sulfate d'ammonium 2M, 5% MPD (2,4-méthyl-pentane-diol). Les données de diffraction ont été collectées sur la ligne ID-23.1 du synchrotron ESRF (European Synchrotron Radiation Facility). 120 images ont été collectées. Le phasage a été effectué par remplacement moléculaire. L'utilisation du programme REFMAC (CCP44) a permis l'affinement de la structure.

### 3.4.7 Spectroscopie RMN

Les expériences RMN ont été enregistrées sur un spectromètre Bruker Avance 700MHz équipé d'une cryosonde. Les données de RMN obtenues ont été traitées et transformées à l'aide du logiciel Topspin (Bruker). Les spectres ont ensuite été analysés et exploités à l'aide du logiciel Sparky (<http://www.cgl.ucsf.edu/home/sparky/>).

Les mesures ont été effectuées à une température de 30°C avec un déplacement chimique de l'eau de 4,79 ppm avec comme référence interne le TSP (3-(triméthylsilyl)[2,2,3,3-2H<sub>4</sub>] propionate). Des échantillons de 500  $\mu$ L de protéine à une concentration de 0,6 mM (0,3mM pour Tox62pT3) ont été préparés dans le tampon phosphate 20 mM pH 5. Sont ajoutés 50 $\mu$ L de deutérium pour régler le lock et 1 $\mu$ L de TSP à 1M pour la référence interne.

#### 3.4.7.1 Expériences pour l'attribution des déplacements chimiques :

Des spectres <sup>1</sup>H <sup>15</sup>N HSQC, <sup>1</sup>H <sup>1</sup>H COSY, <sup>1</sup>H <sup>1</sup>H TOCSY et <sup>1</sup>H <sup>1</sup>H NOESY ont été enregistrés pour les toxines Tox61, Tox62, Tox61pT3, Tox62pT3. Afin de faciliter l'attribution des déplacements chimiques des atomes de Tox61pT3 et Tox62pT3, des expériences de <sup>1</sup>H <sup>1</sup>H <sup>15</sup>N TOCSYHSQC 3D et <sup>1</sup>H <sup>1</sup>H <sup>15</sup>N NOESYHSQC 3D ont été enregistrées.



#### 3.4.7.2 Expériences de dynamique

Afin d'obtenir des informations sur l'état de structuration de ces différentes toxines, des spectres de NOE hétéronucléaires ( $^1\text{H} \rightarrow ^{15}\text{N}$ ) ont été enregistrés pour toutes les molécules. Ce type d'expérience fournit une valeur dépendant des mouvements du vecteur azote-proton à une échelle de temps de l'ordre de la picoseconde à la nanoseconde.

Par ailleurs la toxine Tox61pT3 est en équilibre entre deux formes qui s'échangent en un temps compris entre 10 et 100 millisecondes. Afin de caractériser cet échange et de faciliter l'attribution des pics correspondant à la forme partiellement déstructurée de Tox61pT3 une expérience de  $^1\text{H}$   $^{15}\text{N}$  ROESY hors résonance a été effectuée dans des conditions telles que les interactions dipolaires sont annulées. Des pics de corrélation apparaissent entre des atomes passant d'un environnement chimique à l'autre pendant le temps de mélange de l'expérience à savoir 100 ms (voir Figure 41 en partie 1.3.5 des résultats). Les mouvements mis en évidence par cette technique ont lieu à une échelle de temps comprise entre 10 et 100ms.

---

## 4 PRODUCTION ET CARACTERISATION DES ANTICORPS

---

### 4.1 Production des anticorps

#### 4.1.1 Production des peptides épitopes

Les peptides épitopes ont été produits par synthèse chimique par Alain Lecoq (SIMOPRO CEA-Saclay). La stratégie Fmoc/tBu a été mise en œuvre sur un appareil de synthèse Applied Biosystems 433 selon le protocole établi par le constructeur.

Toutes les étapes suivantes de la production des anticorps ont été réalisées par l'équipe de Laurent Bellanger (CEA-Marcoules) ou par la société Proteogenix.

#### 4.1.2 Fusion à une protéine carrier

##### 4.1.2.1 Anticorps fournis par Marcoules

Les peptides ont été fusionnés à de la thyroglobuline bovine (Sigma Aldrich, France) en utilisant du glutaraldehyde ou du N-(3Diméthylaminopropyl)-N'-éthylcarbodiimide en fonction de leur composition en acides aminés.

##### 4.1.2.2 Anticorps fournis par la société Protéogenix

Les peptides ont été fusionnés à l'immunostimulant KLH (Keyhole limpet hemocyanin).

#### 4.1.3 Immunisation des lapins

##### 4.1.3.1 Anticorps fournis par Marcoules :

Des lapins blancs New-Zealand (Laboratoires Charles River) ont été immunisés par cinq injections sous-cutanées, à 28 jours d'intervalle, en utilisant 400 µg d'antigène (les peptides couplés à la thyroglobuline) repris dans l'adjuvant complet ou incomplet de Freund. Des prélèvements ont été effectués 14 jours après chaque injection. Les titres de chacun de ces prélèvements ont été obtenus en ELISA : des solutions de peptide à 5 µg/ml dans un tampon carbonate (50 mM, pH 9) ont été adsorbés sur des microplaques(Nunc) pendant la nuit. Après saturation avec une solution de sérumalbumine bovine à 1% dans du tampon PBS et lavage, différentes dilutions d'antisérums dans du PBS 50mM pH 7,4 sont incubés pour la nuit. La liaison est révélée en utilisant un anticorps anti IgG de lapin conjugué à la peroxydase (Sigma–Aldrich, France) et de la 3,3',5,5'-tetraméthyl-benzidine (Sigma–Aldrich). Afin d'obtenir des solutions d'immunoglobulines G, les sérums ont été précipités à l'acide caprylique.

##### 4.1.3.2 Anticorps fournis par la société Proteogenix

Des lapins blancs New-Zealand (Laboratoires Charles River) ont été immunisés par une injection intradermique et six injections sous-cutanées, selon un protocole durant 90 jours. Les lapins ont ensuite été sacrifiés. 40 à 50mL de sérum ont été obtenus puis testés en ELISA.

## 4.2 Purification et concentration

### 4.2.1 Chromatographie d'échange d'ions

Cette chromatographie est effectuée sur une colonne Resource Q 6mL (Amersham) montée sur un système FPLC Amersham. Deux tampons sont utilisés pour cette chromatographie : un tampon A composé de Tris-HCl 20mM pH8, un tampon identique au tampon A contenant en plus 1M de NaCl. Pour les opérations de préparation de la colonne, le débit du système est fixé à 1mL/min. La colonne est tout d'abord équilibrée en tampon A. L'échantillon d'anticorps brut, préalablement dialysé contre le tampon de chromatographie est chargé sur la colonne. La colonne est lavée avec 20mL de tampon A. Un gradient effectué à 2 mL/min permet de passer de 0 à 0,6M de NaCl en 25 minutes. La fraction obtenue peut ensuite être concentrée.

### 4.2.2 Chromatographie d'affinité sur colonne de protéine A

Cette chromatographie est réalisée sur colonne Hitrap Protein A de 1mL (Amersham). La colonne est intégrée dans un système Biorad « Biologic » composé d'une pompe péristaltique, d'un détecteur UV. La chromatographie est suivie par mesure de l'absorbance à 280nm. Deux tampons sont utilisés : le tampon de liaison composé de tampon phosphate de sodium à 20mM pH 7 et le tampon d'élution : un tampon citrate 100mM pH 3,5. Pour toutes les opérations suivantes, le débit du système est fixé à 1mL/min. La colonne est lavée avec 20 volumes de colonne de tampon de liaison. 20mg d'anticorps (protéine totale de la solution brute) préalablement filtré sur filtre 0,45µm. La colonne est ensuite lavée jusqu'à retour à la ligne de base. Les anticorps sont alors élués avec le tampon d'élution. Des fractions de 1mL sont prélevées. Chaque fraction est neutralisée avec 300µL de Tris 1M pH9. La colonne est ensuite lavée avec le tampon de liaison jusqu'au retour à la ligne de base.

### 4.2.3 Concentration des anticorps purifiés sur tube vivaspin

Les solutions d'anticorps purifiées par chromatographie d'échange d'ions ou d'affinité sont concentrées sur tube vivaspin (vivascience) de 10kD jusqu'à obtenir un concentrat de 250µL environ.

## 4.3 Caractérisation des anticorps

### 4.3.1 Spécificité

#### 4.3.1.1 ELISA (Enzyme Linked ImmunoSorbant Assay) de titrage

- Titre sur polySC-biot immobilisé sur streptavidine

Cette expérience est réalisée dans une plaque Nunc 96 puits maxisorb. L'activation de la plaque est réalisée par ajout de 100 µL dans chaque puits d'une solution de streptavidine à 2 mg/L dans du tampon PBS. La plaque est incubée une nuit à 4°C. Elle est ensuite saturée par ajout de 300 µL d'une solution de gélatine à 1 g/L (contenant du Thimerosal 0,003% afin d'éviter le développement de micro-organismes). Cette solution est incubée 2 heures à 37°C. Le contenu de la plaque est jeté puis la plaque est lavée 4 fois avec une solution de PBS tween 0,1% vol. On réalise l'incubation de l'antigène par ajout dans chaque puits de 100µL de polySC-biot à 2,7 nM (0,02 µg/mL dans PBS tween 0,1%). La plaque est incubée 2h à 37°C. 4 lavages en PBS tween 0,1% sont ensuite effectués. On place dans chaque puits 100µL de solution d'anticorps primaire diluée à la concentration voulue dans du PBS Tween 0,1% vol. La plaque est incubée 2h à 37°C. 4 lavages en PBS Tween 0,1%. 100µL d'une solution d'anticorps de chèvre anti anticorps de lapin

conjugué à la peroxydase (Jackson ImmunoResearch) au 50 000<sup>ème</sup> dans le PBS Tween sont ajoutés dans chaque puits. Incubation 1h à 37°C. 7 lavages en PBS Tween 0,1%. La révélation est effectuée à l'ABTS (2,2'-azino-bis(3)-ethylbenzthiazoline-6-sulphonic acid). On ajoute 150 µL d'une solution de révélation composée de tampon acétate de sodium 0,1M pH 4,3, de 0,56 g/L d'ABTS (1 tablette de 10 mg pour 18 mL de solution), de 0,01% vol de H<sub>2</sub>O<sub>2</sub> (5,4 µL dans 18 mL). On laisse la réaction de révélation se dérouler pendant 20 minutes à 37°C puis la plaque est lue par mesure de l'absorbance à 405 nm.

- Titre sur polySETag directement immobilisé :

Cette fois-ci l'antigène est immobilisé directement sur la plaque. Lors de l'immobilisation de l'antigène 100µL d'une solution à 1µg/mL de polySETag dans un tampon PBS sont distribués dans chaque puits, puis la plaque est incubée pour la nuit à 4°C. Le protocole continue à l'étape de saturation à la gélatine et est identique à ce qui est décrit ci-dessus.

#### 4.3.1.2 ELISA de compétition

- Test contre les peptides et contre polyS

Cette expérience est réalisée dans les mêmes conditions que l'ELISA de titrage. L'objectif est de réaliser une compétition entre un antigène qui nous intéresse et l'antigène fixé sur la plaque. Parallèlement à la phase de liaison de l'antigène biotinylé à la streptavidine immobilisée sur la plaque d'ELISA, l'anticorps étudié (à une concentration fixe correspondant au titre précédemment déterminé) est incubé dans un tube Eppendorf en présence de l'antigène libre testé (en concentration croissante) dans 100µL d'une solution de PBS Tween 0,1%. Ensuite, lors de la phase d'incubation de l'anticorps primaire, la solution d'anticorps primaire est remplacée par le mélange incubé en Eppendorf contenant l'anticorps primaire et l'antigène libre compétiteur. Le reste du protocole est identique à celui utilisé pour le titrage.

- Test contre les toxines réduites et les toxines oxydées

Des mesures particulières sont employées lorsque l'antigène libre testé est une protéine réduite. Une réduction comme décrit en partie 3.1 de ces matériels et méthodes est effectuée. Après dessalage par HPLC, des quantités croissantes de protéine réduite sont distribuées dans des microtubes Eppendorf et lyophilisées. La protéine réduite lyophilisée est directement reprise avec la solution d'anticorps primaire. Dans ce cas la solution d'incubation est un tampon phosphate citrate 100 mM pH6 contenant 10 mM de DTT, afin de prévenir l'oxydation de la toxine étudiée. La compétition avec la protéine oxydée est effectuée dans les mêmes conditions excepté pour la composition du tampon d'incubation qui ne contient pas de DTT. Il a été vérifié au préalable que la présence de DTT à une concentration de 10mM n'a pas d'effet sur la liaison de l'anticorps.

Dans le cas de toutes les expériences de compétition effectuées avec l'anticorps anti pTxETag, l'antigène immobilisé sur la plaque est polySETagC-biot. La suite des protocoles reste identique.

### 4.3.2 Détermination de la concentration en anticorps actifs en Biacore

La technologie Biacore permet de suivre la liaison entre deux protéines en temps réel. Un des partenaires de la liaison est immobilisé sur une puce (fonctionnalisation), un flux d'une solution de l'autre partenaire est mis en contact avec la surface fonctionnalisée. La grande sensibilité de cette méthode est liée à son principe de détection qui utilise le phénomène de résonance plasmonique de surface. Le signal généré (exprimé en RU pour Resonance Unit) est proportionnel à la masse totale de protéines liées à la puce. Les expériences de Biacore décrites dans cette étude ont été effectuées sur un appareil Biacore 3000. Le tampon utilisé (dénommé par la suite tampon Biacore) est composé d'Hepes 10mM pH 7,4, de NaCl 150mM, d'EDTA 3,4mM et de surfactant P20 (produit Biacore) à 0,05 %vol. La régénération des puces (c'est à dire la dissociation totale des anticorps qui se sont liés à l'antigène immobilisé sur la puce) est assurée par l'utilisation d'une solution d'acide formique 0,5 M. Dans les expériences détaillées ci-après, des puces fonctionnalisées commercialement avec de la streptavidine ont été utilisées. L'injection de solutions à 5nM dans le tampon Biacore de polySC-biot et polySETagC-biot a permis de fixer ces antigènes sur ces puces. Tous les sensorgrammes (courbes d'évolution du signal en fonction du temps) obtenus sont corrigés par soustraction du sensorgramme de la cellule de référence qui n'a pas été fonctionnalisée.

#### 4.3.2.1 Principe de la méthode

((Christensen, 1997), (Richalet-Secordel et al., 1997), (Karlsson et al., 1993))

Le mécanisme de liaison d'un anticorps (Ac) sur une toxine immobilisée sur une puce de Biacore (Tx) peut être décrit par les deux étapes suivantes :



Avec  $\text{Ac}_{\text{sol}}$  l'anticorps en solution,  $\text{Ac}_{\text{surf}}$  l'anticorps à la surface de la puce,  $L_m$  le coefficient de transport de masse, Tx la toxine liée à la puce, AcTx le complexe anticorps-toxine,  $k_a$  la constante cinétique d'association et  $k_d$  de dissociation. La première réaction décrit le transport de masse de l'anticorps entre la solution et la surface de la puce et la deuxième la liaison de l'anticorps sur la toxine immobilisée. Une liaison bimoléculaire est supposée ici. A partir de ce mécanisme réactionnel et en modélisant l'écoulement de la solution d'anticorps dans la cellule de mesure, Christensen ((Christensen, 1997)) aboutit à l'équation différentielle suivante :

$$(3) \quad \frac{dR}{dt} = \frac{L_m \cdot (MM \cdot G \cdot L_r \cdot [\text{Ac}_{\text{sol}}] - k_d \cdot R)}{L_m + L_r}$$

Avec  $L_r = k_a [\text{Tx}]$

et R la valeur de la réponse donnée par l'appareil, MM la masse molaire de l'anticorps, G la réponse de l'appareil par unité de masse de protéine et par surface (qui vaut environ 1000 RU mm<sup>2</sup>/ng : (Stenberg et al., 1991)). Lors de la phase initiale d'association l'influence de la dissociation peut être négligée.

L'équation 3 se réduit alors à :

$$(4) \quad \frac{dR}{dt} = \frac{MM \cdot G \cdot L_r \cdot L_m \cdot [Ac_{sol}]}{L_m + L_r}$$

$L_m$  le coefficient de transport de masse peut être calculé grâce à l'équation suivante :

$$(5) \quad L_m = C_{Lm} \cdot \sqrt[3]{\frac{D^2 \cdot F}{h^2 \cdot l \cdot L}}$$

$$\text{Avec : } C_{Lm} = 1,47 \cdot \frac{1 - (l_1/l_2)^{2/3}}{1 - l_1/l_2}$$

et  $D$  le coefficient de diffusion des anticorps dans le solvant (en  $m^2/s$ ),  $F$  le débit volumique de la solution d'anticorps dans la cellule de mesure (en  $m^3/s$ ),  $h$ ,  $l$ ,  $L$  respectivement la hauteur, la largeur et la longueur de la cellule de mesure (en  $m$ ) et enfin  $l_1$  et  $l_2$  les distances (en  $m$ ) entre lesquelles la détection a lieu.

Dans l'équation (4),  $MM$ ,  $G$ ,  $L_m$  sont connus pour l'appareil de mesure utilisé (Biacore 3000) et la protéine étudiée ( $MM_{\text{anticorps}} = 150000$  Da). En réalisant plusieurs expériences de liaison en faisant varier le débit de solution d'anticorps dans la cellule de mesure, il est possible d'obtenir plusieurs valeurs de  $dR/dt$  alors que  $L_r$  et  $[Ac_{sol}]$  sont constants. L'utilisation d'un logiciel de modélisation numérique (en l'occurrence Kaleidagraph) permet de déterminer les valeurs de  $L_r$  et de  $[Ac_{sol}]$  minimisant l'erreur avec le modèle dans ces différentes expériences.

#### 4.3.2.2 Application de la méthode à notre étude

Le but est de se placer dans des conditions dans lesquelles, le transport de masse est limitant. C'est le cas lorsque la vitesse d'association de l'anticorps sur la toxine immobilisée est grande au regard de la vitesse de diffusion de cet anticorps dans la solution. Une puce où une grande quantité de toxine a été immobilisée doit être utilisée, en l'occurrence 540 RU pour polySC-biot et 700RU pour Tox62 ETag.

Pour illustrer l'application de cette méthode, voici l'exemple de la détermination en anticorps actif de la solution brute de acpTx2 (voir partie 2.1 des résultats). Cet anticorps a été dilué 40 fois dans le tampon Biacore. On réalise des expériences de liaison de l'anticorps à différents débits : 1, 2, 5, 10, 25, 50, 100  $\mu L/min$  (Figure 22).

Trois plages de résultats sont considérées séparément : 1 : 20 à 40 RU, 2 : 40 à 60 RU, 3 : 60-80 RU. Pour chaque plage de données et pour chaque débit utilisé,  $dR/dt$  est déterminé par régression linéaire. L'équation (5) permet de calculer la valeur de  $L_m$  pour chaque débit. Pour calculer  $L_m$ , les valeurs suivantes ont été utilisées.  $D$ , le coefficient de diffusion des anticorps dans le solvant vaut  $4.10^{-11} m^2/s$  ((Karlsson et al., 1993)).  $h$  vaut 0,05mm,  $l$  0,5mm et  $L$  2,1mm pour le Biacore 3000 (d'après les données fournies par le constructeur). La surface de liaison est supposée être un rectangle de 1,6 mm de long centré sur le milieu de la cellule de mesure et donc  $l_1$  vaut 0,25mm et  $l_2$  1,85mm ((Richalet-Secordel et al., 1997)).

A partir de ces valeurs de  $L_m$  calculées et des valeurs de  $dR/dt$  déterminées expérimentalement, on peut tracer  $dR/dt = f(L_m)$  (Figure 23).

La valeur de concentration en anticorps actifs retenue correspond à la moyenne des valeurs obtenues pour les différentes plages de résultat considérées. Les valeurs de  $L_r$  peuvent être utilisées pour vérifier que l'hypothèse de limitation de transport de masse est valide.

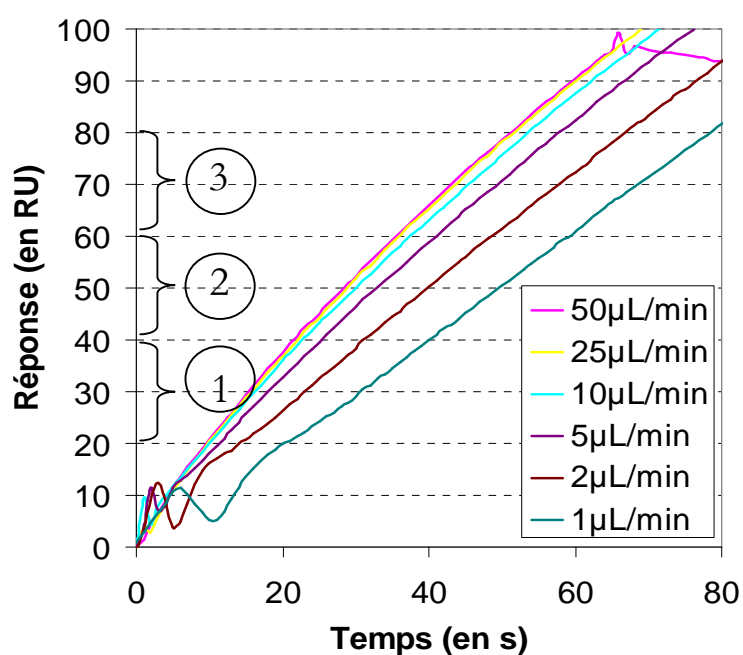


Figure 22: Courbes de réponse obtenues lors de la liaison de acpTx2 brut dilué 40 fois à différents débits, dans le cadre de la détermination de la concentration en anticorps actif.

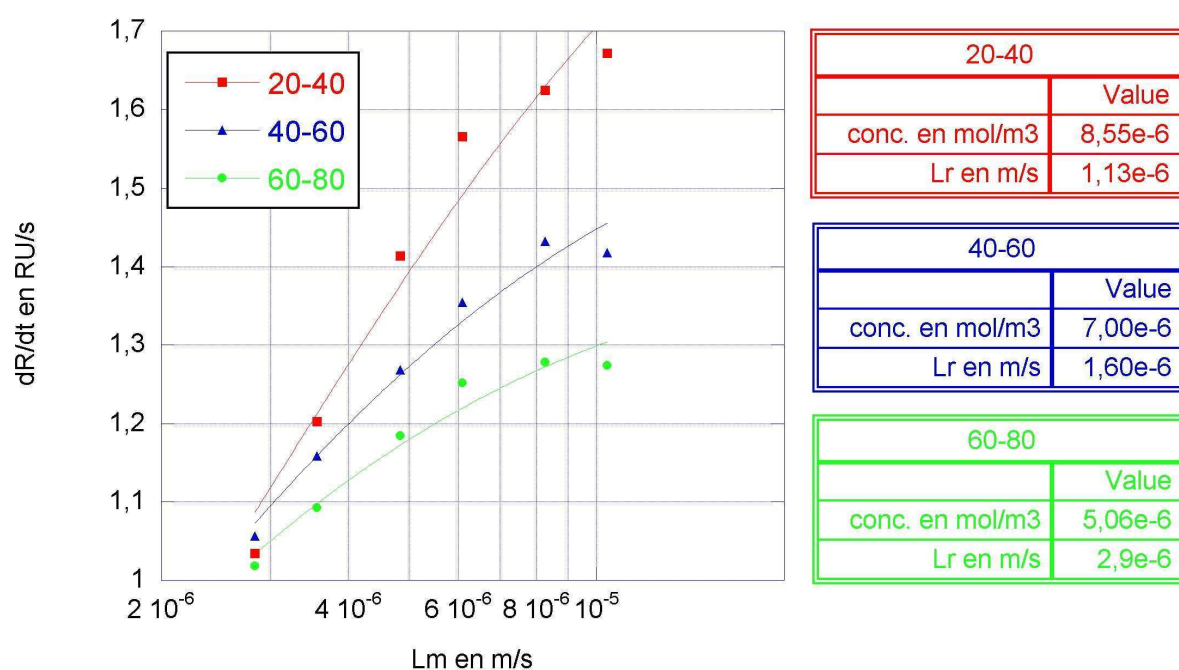


Figure 23: A gauche,  $dR/dt$  déterminés expérimentalement en fonction des  $L_m$  calculés. A droite, les tableaux présentent les résultats de la modélisation effectuée à partir de l'équation (4) pour chacune des plages de résultat considérées.

### 4.3.3 Liaison de deux anticorps sur un antigène

#### 4.3.3.1 Etude en gel filtration et immunoblot :

##### ■ Gel filtration

Les expériences de gel filtration ont été réalisées sur une colonne Superdex 200HR 10/30 (de volume total 25mL) sur un système Äkta. Le chargement et l'élution ont été effectués en tampon PBS à un débit de 0,25 mL/min. Dans une première étape la colonne a été calibrée.

Le  $K_{av}$ , est calculé de la façon suivante :

$$K_{av} = (V_e - V_0) / (V_T - V_0)$$

avec dans ce cas  $V_0 = 6,8\text{mL}$ ,  $V_T = 25\text{mL}$

Afin de calibrer la colonne, un mélange de protéines (ferritine, aldolase, albumine et ovalbumine) a été injecté sur la colonne de gel filtration (Tableau 5).

	Masse moléculaire (en Da)	Volume d'exclusion (en mL)	$K_{av}$
Ferritine	440 000	9,1	0,13
Aldolase	158 000	10,82	0,23
Albumine	67 000	12,35	0,32
Ovalbumine	43 000	13,4	0,39

Tableau 5 : Résultats de la calibration de gel filtration.

Le  $K_{av}$  dépend linéairement du rayon hydrodynamique de la protéine considérée. Sachant que les protéines utilisées pour la calibration ainsi que la toxine et les anticorps étudiés sont globulaires, on peut considérer que le rayon hydrodynamique de ces protéines est proportionnel à leur masse moléculaire. L'équation suivante permet d'obtenir  $K_{av}$  en fonction de la masse moléculaire de la protéine étudiée dans notre système :

$$K_{av} = -0,24 \times \log_{10}(MM) + 1,49$$

Grâce à cette équation et à la formule de  $K_{av}$ , il est possible de déterminer la masse moléculaire d'une espèce à partir de son volume d'élution et vice-versa.

##### ■ Immunoblot

Les fractions analysées ont été séparées en SDS-PAGE sur gel de polyacrylamide (14%) en tampon Tris/Tricine. Le gel obtenu a été transféré sur membrane de PVDF de 0,4 $\mu\text{m}$  (Immobilon). Le tampon cathode de transfert est composé de tampon Tris 25mM, Glycine 40mM pH9,4 avec 10% de méthanol. Les tampons anode I et II contiennent du Tris-HCl pH 10,2 à 300mM et 25mM respectivement et avec 10% de méthanol. La migration est effectuée en courant constant à 4mA/cm<sup>2</sup> pendant 30 minutes. La membrane est ensuite saturée pendant 1 heure dans une solution de BSA 2% dans un tampon PBS. Elle est ensuite incubée pendant 1 heure dans une solution d'anticorps acpTx2 et acpTx3 (voir partie 2.1 des résultats) au 100<sup>ème</sup> dans du tampon PBS, Tween 0,05%, BSA 1% puis pendant 1 heure dans une solution de v (Goat Anti-Rabbit PerOxydase) diluée au 2000<sup>ème</sup> dans du tampon PBS Tween 0,05% BSA 1%. La révélation est effectuée avec une solution de DAB (Diaminobenzidine, Sigma) à 0,5mg/mL dans un tampon Tris-HCl 0,M pH7,6 et contenant 0,2% d'H<sub>2</sub>O<sub>2</sub>.



#### 4.3.3.2 Etude en Biacore :

(Voir la partie 4.3.2 de ces matériels et méthodes pour la méthodologie générale liée à l'utilisation du Biacore.)

L'objectif de cette expérience est de s'approcher des conditions de saturation de l'antigène avec un anticorps puis d'observer que lorsqu'on injecte le deuxième anticorps la liaison a lieu, de la même façon que si aucun antigène n'avait été préalablement lié. Une puce fonctionnalisée avec une faible quantité d'antigène, en l'occurrence 40 RU de polySC-biot, est utilisée ici. Les anticorps sont dilués 40 fois dans du tampon Biacore. Le débit est fixé à 5µL/min. L'expérience type comporte 4 étapes :

- injection de 650µL du premier anticorps
- injection de 650µL du deuxième anticorps
- injection de 650µL du peptide épitope du deuxième anticorps injecté
- injection de 650µL du peptide épitope du premier anticorps injecté

#### 4.3.4 *Détermination de la vitesse de dissociation de la toxine réduite*

La détermination de la vitesse de dissociation des anticorps a été effectuée par radio immuno assay (RIA). L'objectif de cette expérience est d'incuber l'anticorps étudié avec la toxine réduite marquée au tritium en conditions réductrices (pour maintenir la toxine réduite). Au temps zéro, on ajoute un excès de peptide épitope, on suit en fonction du temps la quantité de complexe anticorps/toxine réduite restant.

##### 4.3.4.1 Réduction de la toxine tritiée et dessalage

La réduction de 2 nmol de Tox62 <sup>3</sup>H à 920 Ci/mol est effectuée dans un volume de 500µL. Elle est menée dans les mêmes conditions que décrit en partie 3.1 de ces matériels et méthodes. Le dessalage de la réaction de réduction est effectué sur colonne de gel filtration PD10 en condition acide, de la façon suivante. La colonne est équilibrée avec 25mL d'acide acétique 0,5%. Le mélange de réduction est dilué dans 2mL d'acide acétique 0,5%. Le mélange obtenu est chargé sur la colonne. La protéine dessalée est éluée avec 3,5 mL d'acide acétique 0,5%. L'éluat est lyophilisé pendant la nuit.

##### 4.3.4.2 Déroulement de l'expérience de cinétique de dissociation

La toxine réduite et dénaturée est reprise dans 500µL d'acide acétique 0,5%. 10µL de cette solution sont comptés afin d'établir la concentration en toxine réduite dans la solution. Le mélange d'incubation est composé de tampon Tris-HCl à 100mM pH 7,6, de l'anticorps étudié à 10nM d'anticorps spécifiques, de TCEP à 8µM et de toxine Tox62 <sup>3</sup>H réduite à 100nM. 2 mL de ce mélange sont incubés 2h à 25°C. Au temps zéro, on ajoute le peptide épitope pour atteindre une concentration de 10µM soit un excès 100 fois par rapport à la concentration en anticorps spécifiques. A différents temps (0', 5', 10', 20', 40'), 300µL de prélèvement sont ajoutés à 20µL de sérum normal de cheval. Le tube est placé dans la glace pendant 1minute 30 secondes. 500µL de PEG 6000 (Polyéthylène glycol 6000 Da) à 20% (masse de PEG/masse d'eau) sont ajoutés. Ce temps de 1 minute 30 est ajouté aux temps de prélèvement. Ainsi le point 0 minutes devient le temps 1minute 30, le temps 5 minutes le temps 6 minutes 30 etc... Le mélange est homogénéisé puis centrifugé à 10 000 trs/min à 4°C pendant 5'. Le surnageant est éliminé et le culot est repris dans 750 µL de soude 0,05N. Cette solution est mélangée à 10mL de scintillant liquide Aqualuma (Perkin Elmer) dans une fiole de comptage. La fiole est comptée.

#### 4.3.4.3 Calcul du pourcentage de liaison :

Trois expériences de dissociation ont été menées parallèlement, les deux premières testaient la dissociation de acpTx2 et acpTx3 ; dans la troisième, les anticorps étaient remplacés par des anticorps de lapin non spécifiques à 0,5mg/mL dans la solution d'incubation (ce qui correspond à la concentration en anticorps totaux pour acpTx3). Pour les deux premières expériences, un point est prélevé après les deux heures d'incubation, avant la chasse par le peptide, ce point représente le maximum lié (noté Lmax). Les cinq points obtenus dans la troisième expérience donnent le signal non-spécifique (Lns). Chaque prélèvement lors des cinétiques de acpTx2 et acpTx3 donnent une mesure de la quantité de complexe notée L. Le pourcentage de liaison (%L) est donné par l'équation suivante :

$$\%L = (L - Lns) * 100 / (Lmax - Lns)$$

Les barres d'erreurs données sur la Figure 48, correspondent à l'écart type de répétabilité des cinq mesures de liaison totale.

## 5 EFFET DES ANTICORPS SUR LE REPLIEMENT

### 5.1 Expérience de repliement suivi par mesure de l'activité biologique

#### 5.1.1 Mesure de l' $IC_{50}$ d'une toxine

Les neurotoxines ont la capacité de se lier aux récepteurs musculaires nicotiniques à l'acétylcholine et notamment aux récepteurs nicotiniques à l'acétylcholine présents en abondance dans l'organe électrique de torpille (*Torpedo marmorata*). Afin de tester l'affinité de Tox62, on réalise une expérience de compétition de liaison sur des préparations de membranes de cellules d'organe électrique de torpille ((Tremeau et al., 1995)) entre Tox62 et une neurotoxine marquée (bungarotoxine iodée 125 : Perkin Elmer NEX126 : 120Ci/mmol) avec une concentration en Bungarotoxine marquée à l'iode 125 proche de son  $K_d$  pour qu'il soit possible, avec des quantités raisonnables de toxine de remplacer la bungarotoxine iodée par de la toxine  $\alpha$  (Figure 24).

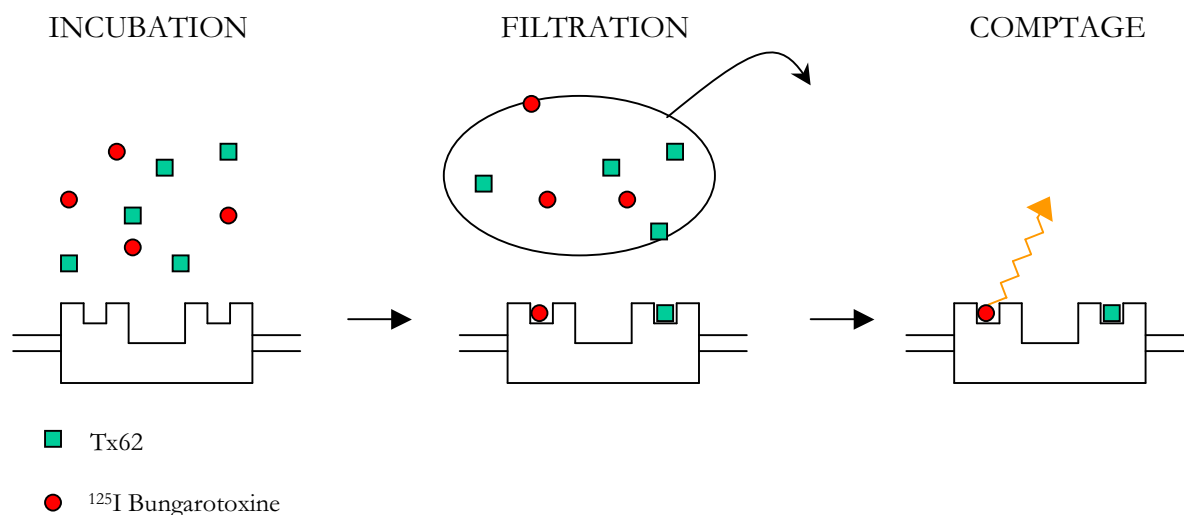


Figure 24 : Schéma du principe du test d'activité biologique

Dans une plaque 96 puits, on réalise l'incubation de différentes quantités de Tox62 avec une quantité fixe de bungarotoxine iodée et de récepteurs de torpille. L'incubation est effectuée dans un tampon composé de tampon PBS, de BSA (Bovin Serum Albumin) à 0,1% massique, de récepteurs de torpille à 0,2nM, de Bungarotoxine iodée à 0,4nM et de Tox62 oxydée à des concentrations variant de 10pM à 100nM. Pour chaque concentration trois points sont effectués. La plaque ainsi préparée est incubée pour la nuit à température ambiante. Le lendemain une plaque de filtres est préparée sur lesquels ont été préalablement déposés 40 $\mu$ L de PEI (Polyéthylèneimine) à 0,1%volume dans l'eau. Les solutions contenues dans les puits de la plaque sont filtrées sur la plaque de filtres grâce à un système de filtration de plaques. Les filtres sont rincés par trois fois avec 300 $\mu$ L de PBS. La plaque de filtres est séchée. 25 $\mu$ L de scintillant est ajouté sur chacun des filtres et la plaque est comptée.

La courbe de compétition obtenue (S) est modélisée par l'équation suivante :

$$S = S_{ns} + (S_{max} - S_{ns}) / (1 + ([Tox62] / IC_{50}))$$

Avec S le signal observé,  $S_{max}$  le signal maximum (quand il n'y a pas de compétiteur),  $S_{ns}$  le signal non spécifique (mesuré avec un très large excès de compétiteur connu),  $IC_{50}$  la valeur de la concentration en Tox62 permettant d'obtenir 50% de liaison et  $[Tox62]$  la concentration en Tox62 dans le puits de la plaque. La modélisation des courbes de compétition est effectuée grâce au logiciel Kaléidagraph pour déterminer la valeur de l' $IC_{50}$  de Tox62 ou d'autres toxines.

L'équation peut être inversée pour déterminer la concentration en Tox62 dans le puits à partir du signal observé S et de la connaissance de l' $IC_{50}$  de Tox62 dans les conditions de l'étude :

$$[Tox62] = IC_{50} * ((S_{max} - S_{ns}) / (S - S_{ns}) - 1)$$

### 5.1.2 Expérience de repliement en mesure d'activité

Le protocole décrit en partie 3.1 de ces matériels et méthodes est utilisé pour réduire 50µg de Tox62. Le repliement est effectué dans un volume de 1,5mL de tampon de repliement composé de tampon Tris-HCl 0,1 M pH 7,5, d'EDTA à 1mM, de GSH à 3 mM, de GSSG à 0,3 mM GSSG, de Tox62 réduite à 2nM (ajoutée avec un cône silanisé – REPEL Silan : GE Healthcare - pour éviter les pertes sur les parois) et enfin d'immunoglobulines G non spécifiques de lapin à 0,7mg/mL. La solution de blocage contient du iodoacétamide à 0,3M, du peptide pTx2 à 15µM, du peptide pTx3 à 15µM. Au temps zéro, la toxine réduite est ajoutée au tampon de repliement. Le repliement est effectué à 25°C. A différents temps (0', 30', 1h30, 3h, 6h, 24h), 240µL du milieu réactionnel sont prélevés, auxquels sont ajoutés 120µL de solution de blocage (ce qui donne une concentration finale en iodoacétamide de 0,1M et de 5µM en peptides). Les échantillons bloqués sont laissés une heure à 25°C. Ils sont ensuite congelés à l'azote puis conservés à -20°C.

Les échantillons obtenus sont décongelés au bain marie pendant 5' à 25°C. Chaque échantillon de 360µL est divisé dans trois puits en trois fois 100µL. Ces 100µL sont analysés exactement comme décrit précédemment pour l'obtention de l' $IC_{50}$ . Une courbe de compétition réalisée en présence du tampon de repliement et de la solution de blocage a permis de vérifier que l' $IC_{50}$  de Tox62 n'est pas modifiée dans ces conditions.

### 5.1.3 Inhibition de la réaction de repliement par les anticorps suivie en activité biologique

La même réaction de repliement que décrit ci-dessus est réalisée. Cette fois-ci les anticorps non spécifiques sont remplacés par des anticorps spécifiques à une concentration de 10nM en anticorps spécifiques soit un excès 5 fois par rapport à la concentration en toxine réduite. En fonction de la concentration totale en IgG de ces solutions d'anticorps spécifiques, la concentration en anticorps totale est ajustée par l'ajout d'immunoglobulines G non spécifiques de lapin (Sigma).

### 5.1.4 Levée de l'inhibition par ajout des peptides :

La même réaction que décrit ci-dessus est réalisée. Dans ce cas, une première phase permet l'incubation de la toxine réduite avec les anticorps spécifiques en condition réductrice puis dans un deuxième temps l'oxydant et le peptide épitope sont ajoutés. Cette expérience doit permettre de vérifier que l'inhibition du repliement par les anticorps est réversible. Une solution de repliement comme décrit ci-dessus pour l'étude de l'inhibition est préparée à ceci près que cette solution ne contient pas de glutathion oxydé. Après cinq minutes d'incubation on ajoute une solution contenant le glutathion oxydé (pour obtenir une concentration finale de 0,3mM) et le peptide épitope correspondant à l'anticorps pour atteindre une concentration de 3μM soit 300 excès par rapport aux anticorps. Le reste de l'expérience est réalisé comme dans le cas des expériences d'inhibition.

## 5.2 Expérience de repliement suivi en isoélectrophorèse

Pour ces expériences, la toxine Tox62 Etag <sup>35</sup>S a été utilisée.

1μg de Tox62Etag <sup>35</sup>S à 14000Ci/mol sont dénaturés dans 500μL dans les conditions décrites en partie 3.1 de ces matériels et méthodes. Le repliement de Tox62Etag à une concentration de 20nM est effectué dans des conditions identiques à celles décrites en partie 3.1 de ces matériels et méthodes, seul le tampon de repliement a été modifié car il contient des immunoglobulines G à une concentration de 1mg/mL, et éventuellement des anticorps actifs à une concentration de 100nM. Enfin, la réaction est menée dans un tube minisorb afin de limiter les pertes de toxine réduite sur les parois du tube. Des échantillons de 20μL sont prélevés aux temps 0', 30', 1h, 2h, 4h, 24h. 1μL de solution de MTSBS 1M (voir partie 3.2.2 de l'introduction) dans le DMSO (Dimethyl sulfoxide) et 1μL d'une solution du peptide à 300μM correspondant à l'anticorps utilisé (ou d'eau dans le cas des expériences témoin où des anticorps non spécifiques sont utilisés) sont ajoutés à l'échantillon. Toutes les manipulations de la solution de repliement sont effectuées avec des cônes silanisés pour éviter les pertes sur leurs parois. Le prélèvement est ensuite incubé 10 minutes à 50°C puis plongé dans de l'azote liquide et conservé à -20°C. Ils sont ensuite analysés en isoélectrophorèse. Pour cette analyse le système d'isoélectrophorèse multiphore II (GE Healthcare) est utilisé. Les gels sont des ampholine PAGplate large scale (GE Healthcare). Les échantillons sont décongelés par incubation 5 minutes au bain marie à 25°C. La migration est effectuée pendant 1h 30 à 1500 V, 50 mA pour un gel entier (et proportionnellement moins si le gel est plus petit) et 12°C. Le gel est ensuite désolidarisé de son support puis séché entre deux feuilles de cellophane sur un sécheur à vide à 90°C (le but est de sécher rapidement le gel pour que la toxine n'ait pas le temps de diffuser lorsqu'elle n'est plus focalisée à son point isoélectrique par le champ électrique). Le gel sec est ensuite exposé sur une cassette Storage phosphor screen (Molecular Dynamics). La révélation est effectuée par lecture de la cassette sur un appareil Typhoon (GE Healthcare).

## **RESULTATS**







---

## 1 LE MECANISME DE REPLIEMENT *IN VITRO* DE LA TOXINE $\alpha$

---

### 1.1 La longueur de lk1 influe sur les étapes finales du repliement

Les travaux de Ruoppolo *et al.* ((Ruoppolo et al., 1998)) grâce à l'étude de la vitesse de repliement des variants Tox60, Tox61 et Tox62, ont montré que cette vitesse dépend du nombre de résidus présents dans la partie de la séquence appelée lk1, reliant la boucle 1 à la boucle 2 (Figure 11 en partie 5.3.1 de l'introduction). Pour expliquer ce résultat, les auteurs ont proposé que lors des étapes précoces du repliement la formation des premiers ponts disulfure s'effectue de manière aléatoire. L'espace conformationnel exploré au hasard par la molécule pendant ces étapes serait d'autant plus vaste que cette séquence lk1 serait longue. Par conséquent, les intermédiaires non-natifs précoces s'accumuleraient ce qui expliquerait la lenteur de la cinétique de repliement. Afin de tester cette hypothèse, nous avons mis en œuvre deux méthodes chromatographiques différentes pour étudier la composition en intermédiaires des mélanges de repliement de Tox60, Tox61 et Tox62 dans les étapes précoces puis aux temps longs.

#### 1.1.1 Les étapes initiales du repliement sont similaires entre Tox60, Tox61 et Tox62

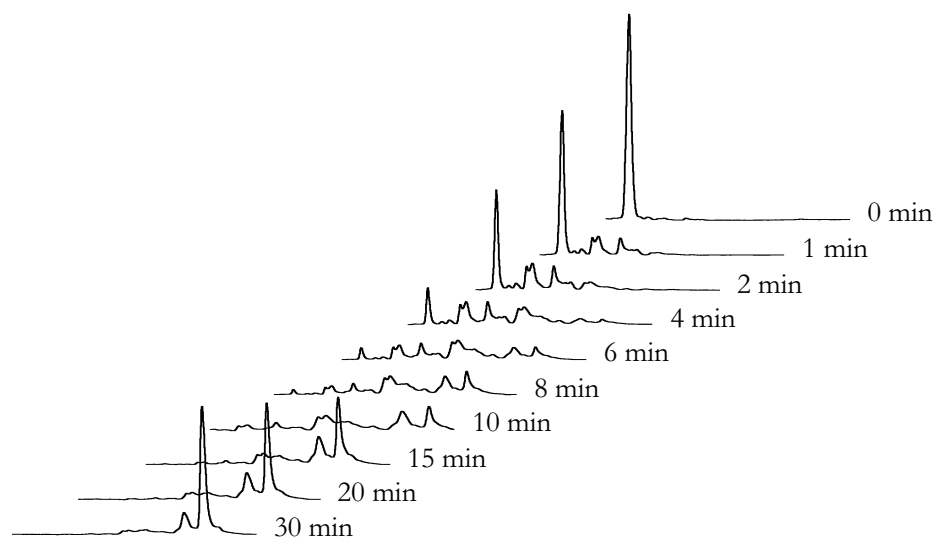
L'objectif de cette partie est de comparer les espèces formées et leur évolution lors des premières étapes du repliement de Tox60, Tox61 et Tox62.

##### 1.1.1.1 Evolution des profils chromatographiques d'échange d'ions des mélanges de repliement

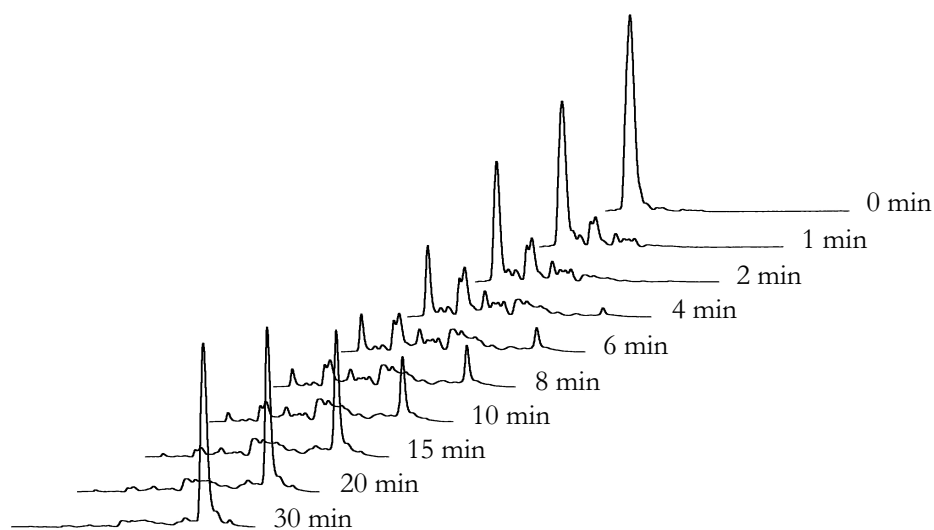
Le principe de ce type d'expérience est de réduire et dénaturer une protéine puis de la placer dans des conditions permettant son oxydation et son repliement. Les premiers intermédiaires à se former lors du repliement d'une protéine à ponts disulfure contiennent un faible nombre de ponts et sont peu structurés. Pour faciliter leur séparation, on utilise une méthode de blocage (ici le MTSBS) empêchant la formation des ponts disulfure et ajoutant une charge négative à chaque cystéine libre ce qui permet de séparer par chromatographie d'échange d'ions les espèces protéiques en fonction de leur nombre de ponts.

Pour ces trois variants, au temps zéro, une seule forme, qui correspond à la toxine réduite, est présente dans le milieu de repliement (Figure 25). Cette forme présente le temps de rétention le plus faible parmi toutes les formes apparaissant par la suite. Dans les trois cas, des intermédiaires présentant des temps de rétention de plus en plus importants apparaissent progressivement pour que finalement deux (une seule dans le cas de Tox61) espèces disposant du temps de rétention le plus important s'accumulent. L'évolution des profils chromatographiques pour les différents mutants étudiés ne semble pas modifiée de façon importante dans ces premières étapes (Figure 25). La différence principale est qu'après 15 minutes de repliement il apparaît deux formes majoritaires pour Tox60 et Tox62, alors que pour Tox61 une seule forme s'accumule.

1



2



3

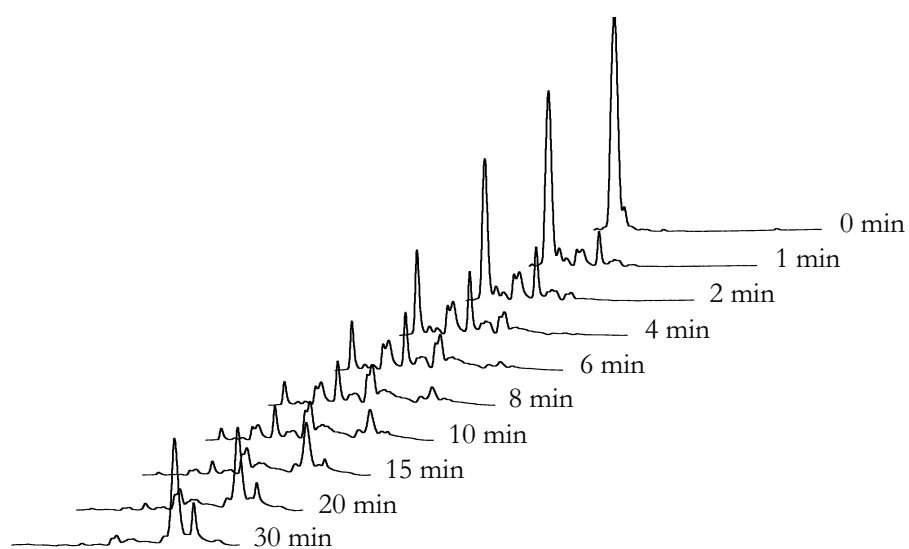


Figure 25 : Repliement oxydant de Tox60 (1), Tox61 (2) et Tox62 (3) aux temps courts, suivi en chromatographie d'échange d'ions

### 1.1.1.2 Analyse des chromatogrammes obtenus après 8 minutes de repliement

Pour simplifier la comparaison, prenons les chromatogrammes obtenus après 8 minutes de repliement (Figure 26). La comparaison de ces profils chromatographiques au niveau des pics 1 à 7 indique qu'ils sont similaires. Le nombre d'intermédiaires formés après 8 minutes n'est pas différent d'un variant à l'autre. La forme réduite (pic 1) de Tox60 a disparu plus rapidement que les formes réduites de Tox61 et Tox62. D'autre part, le pic 3 de Tox62 est plus abondant que pour les autres mutants. L'importance et la forme des pics 8 et 9 diffèrent significativement.

### 1.1.1.3 Analyse de masse des formes obtenues

Les pics 1 à 9 séparés en échange d'ions ont été analysés en spectrométrie de masse.

Pic n°	Tox 60			Tox 61			Tox62		
	Masse	MTSBS	Nbre ponts	Masse	MTSBS	Nbre ponts	Masse	MTSBS	Nbre ponts
1	8020	8	0	8140	8	0	8225,7	8	0
2	7673	6	1	7800	6	1	7890,6	6	1
3	7672	6	1	7806	6	1	7890,1	6	1
4	7334	4	2	7800	6	1	-	-	-
5	7341	4	2	7468	4	2	7551,8	4	2
6	7341	4	2	7467	4	2	-	-	-
7	7341	4	2	-	-	-	-	-	-
8	6995	2	3	7124	2	3	7212	2	3
9	6996+6657	2+0	3+4	-	-	-	7212+6873	2+0	3+4

Tableau 6 : Analyse de masse des intermédiaires purifiés lors des expériences de repliement aux temps courts. Dans la première colonne figure les masses moyennes obtenues. La deuxième colonne indique le nombre de molécules de MTSBS incorporées correspondant à cette masse.

Pour tous les variants il est observé que : le pic 1 correspond à la forme réduite, les pics 2 et 3 à des formes à 1 pont, les pics 5, 6 et 7 à des formes à deux ponts, le pic 8 à une forme à trois ponts et le pic 9 à des formes à 3 ou 4 ponts. La situation du pic 4 est plus ambiguë, c'est une forme à 2 ponts pour Tox60, 1 pont pour Tox61 et l'information n'a pu être obtenue pour Tox62. Ces résultats montrent qu'excepté pour le pic 4, les pics ayant le même temps de rétention ont des nombres de cystéines libres équivalents d'un variant à l'autre. Le pic 9 est le seul pic contenant des espèces ayant des nombres de ponts différents. La méthode utilisée ici de chromatographie d'échange d'ions n'est pas adaptée pour séparer les intermédiaires à trois ponts de la forme native.

On remarque que pour le pic 9 de Tox61, aucune masse n'a pu être obtenue alors que ce pic est grand. Il est probable que, lors de l'étape de dessalage de l'échantillon pour envoi en masse, la protéine ait été perdue.

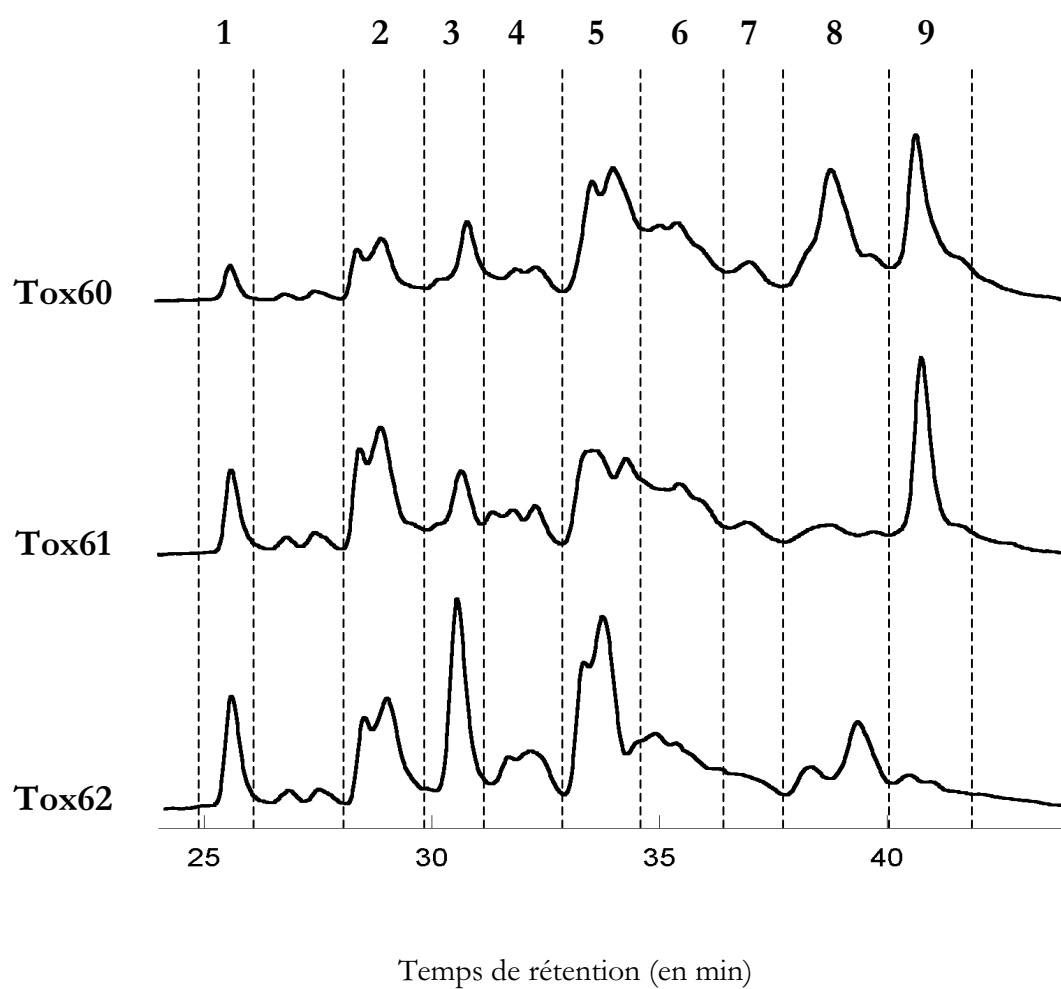


Figure 26 : Profils chromatographiques d'échange d'ions d'échantillons après 8 minutes de repliement.

#### 1.1.1.4 Intégration et courbes cinétiques

A partir des données de masse présentées dans le Tableau 6, les chromatogrammes d'HPLC d'échange d'ions ont été intégrés et les fractions correspondant aux intermédiaires contenant le même nombre de ponts ont été regroupées. Les résultats sont exprimés en pourcentage de l'aire totale en fonction du temps. Comme il n'est pas possible, par cette méthode chromatographique, de séparer certains intermédiaires à 3 ponts de la forme oxydée, les pics correspondant aux formes à 3 et 4 ponts ont été regroupés. Ne disposant pas de masse pour certains pics car leur quantité est faible (Tableau 6), un nombre de ponts égal à celui de la forme ayant le même temps de rétention dans les autres variants leur est attribué (cette approximation semble raisonnable car ces formes apparaissent en faible quantité donc si une erreur est commise elle n'influencera que faiblement le résultat global). Le pic 4 correspond à une forme à 2 ponts dans le cas de Tox60 et à 1 pont dans le cas de Tox61. Pour Tox62 le pic 4 a été classé dans les formes à 1 pont. Les courbes obtenues sont présentées en Figure 27.

Pour les trois variants étudiés, la forme réduite disparaît à la même vitesse, soit un temps de demi-vie d'environ 2 minutes. Les intermédiaires à 1 et 2 ponts évoluent de manière similaire entre Tox61, et Tox62. Dans le cas de Tox60, il semble que les intermédiaires à 1 pont disparaissent plus rapidement pour former des intermédiaires à 2 ponts que dans le cas de Tox61 et Tox62.

50% de formes à 3 ou 4 ponts sont apparus après 10 minutes pour Tox60, 14 minutes pour Tox61 et 16 minutes pour Tox62.

Les formes à 3 ou 4 ponts apparaissent 1,4 fois plus vite pour Tox60 que pour Tox61. Ceci est à rapprocher du fait que Tox60 se replie 1,7 fois plus vite que Tox61 (Tableau 3 en partie 6.1 de l'introduction) et laisse supposer que l'augmentation de la vitesse de conversion des intermédiaires précoces est partiellement responsable de la l'accélération globale du repliement de Tox60

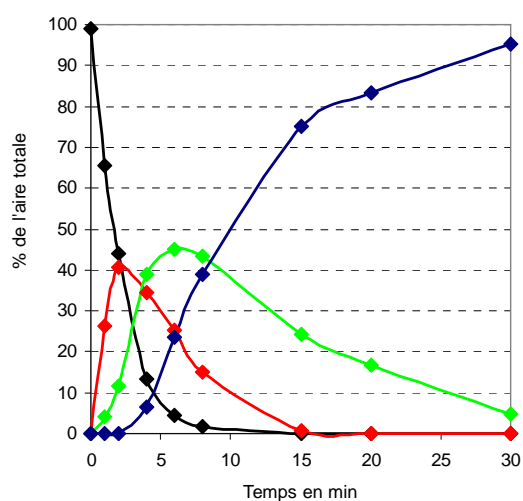
Pour Tox62 les formes à 3 ou 4 ponts apparaissent 1,1 fois plus lentement que pour Tox61. Le rapport de leurs vitesses d'apparition de forme native est de 3,7 (Tableau 3 en partie 6.1 de l'introduction). Le ralentissement de la vitesse de repliement entre Tox62 et Tox61 n'est pas dû à un ralentissement des étapes précoces.

#### 1.1.1.5 Conclusion

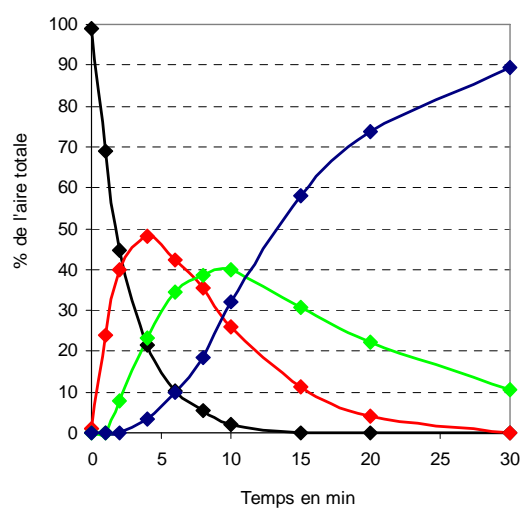
En conclusion, cette analyse a permis de montrer qu'il n'existe pas de différences dans le nombre d'intermédiaires apparaissant dans les premières étapes du repliement de Tox60, Tox61 et Tox62 et donc de rejeter l'hypothèse proposée par Ruoppolo *et al.* ((Ruoppolo et al., 1998))

La comparaison des vitesses d'apparition des formes à 3 et 4 indique que l'essentiel des différences de cinétique d'apparition des formes natives observées n'est pas lié aux étapes précoces du repliement mais plutôt à la vitesse de conversion de ces espèces en forme native.

1



2



3

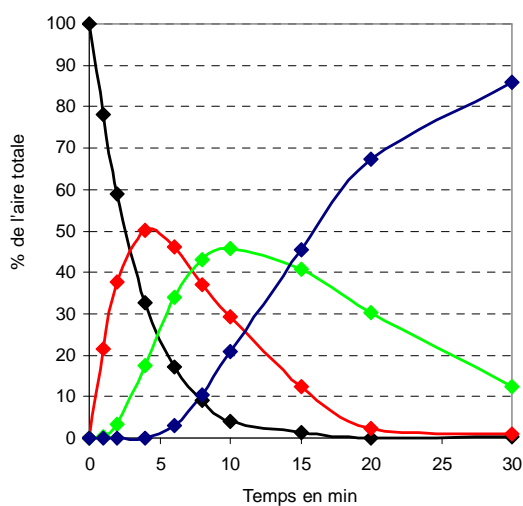


Figure 27 : Cinétiques de repliement aux temps courts de Tox60 (1), Tox61 (2), Tox62 (3) suivies en HPLC d'échange d'ions. En noir : forme réduite, en rouge : formes à 1 pont, en vert : formes à 2 ponts, en bleu : formes à 3 et 4 ponts. Les courbes de cinétique ont été tracées en reliant les points expérimentaux.

### 1.1.2 La longueur de lk1 modifie les étapes finales du repliement

L'objectif est maintenant de détailler l'évolution des espèces à trois et quatre ponts et leur conversion en forme native.

#### 1.1.2.1 Evolution des profils chromatographiques en phase inverse des mélanges de repliement

La méthode de blocage au MTSBS suivi d'une chromatographie d'échange d'ions s'est révélée insuffisante pour séparer les espèces à 3 et 4 ponts. Ici le blocage de la réaction est assuré par une acidification du milieu de repliement, les espèces formées sont séparées en HPLC en phase inverse, donc en fonction de leur hydrophobie de surface. Par conséquent, la forme dénaturée et réduite a un temps de rétention supérieur à la forme oxydée repliée. Les intermédiaires n'ayant que partiellement acquis leur structure auront un temps de rétention compris entre celui de la forme réduite et celui de la forme oxydée. Les chromatogrammes de la Figure 28 présentent le repliement en conditions oxydantes des trois toxines considérées sur une échelle de temps comparable à celle utilisée lors des travaux de Ruoppolo *et al.* ((Ruoppolo et al., 2001)).

Au départ les mélanges de repliement contiennent uniquement la toxine réduite (Figure 28). Après 24 heures, dans les cas de Tox60 et Tox61, 100% de forme oxydée sont obtenus. Pour Tox62, il reste une quantité minime d'intermédiaire. Ces réactions d'oxydation, dans les conditions étudiées sont donc totales. De plus, ces profils nous montrent que pour ces trois variants de la toxine  $\alpha$ , la forme réduite disparaît rapidement, en moins de 30 minutes, pour donner deux intermédiaires qui s'accumulent. On appelle ces intermédiaires C et D dans un souci de cohérence avec la notation proposée par Ruoppolo *et al.* ((Ruoppolo et al., 2001)). D'après leurs travaux, dans le cas de Tox62, C correspond à l'intermédiaire des [43-54] et D à l'intermédiaire des [17-41]. La transformation de ces intermédiaires en la forme oxydée semble constituer l'étape finale du repliement.

#### 1.1.2.2 Intégration et courbes cinétiques

Les chromatogrammes sont utilisés pour réaliser l'intégration des pics correspondant aux différentes espèces apparaissant lors du repliement oxydant de Tox60, Tox61 et Tox62.

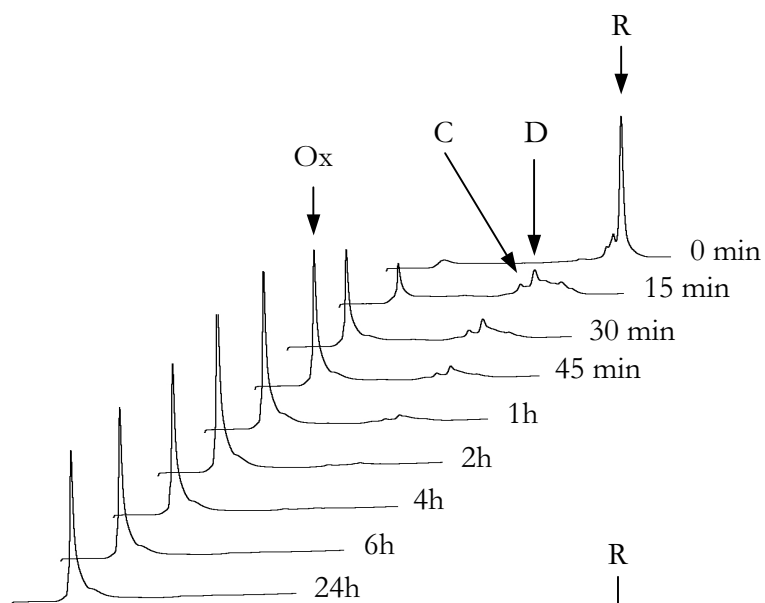
Ces courbes, présentées sur la Figure 29, illustrent les différences de cinétique de repliement existant entre ces variants de la toxine  $\alpha$ . Des temps de demi-repliement de 25, 35 et 230 minutes sont obtenus pour Tox60, Tox61 et Tox62 respectivement. La vitesse de disparition de la forme réduite semble peu affectée.

La quantité d'intermédiaires C et D est quant à elle très différente d'un variant à l'autre. Moins les intermédiaires C et D s'accumulent, plus le repliement est rapide.

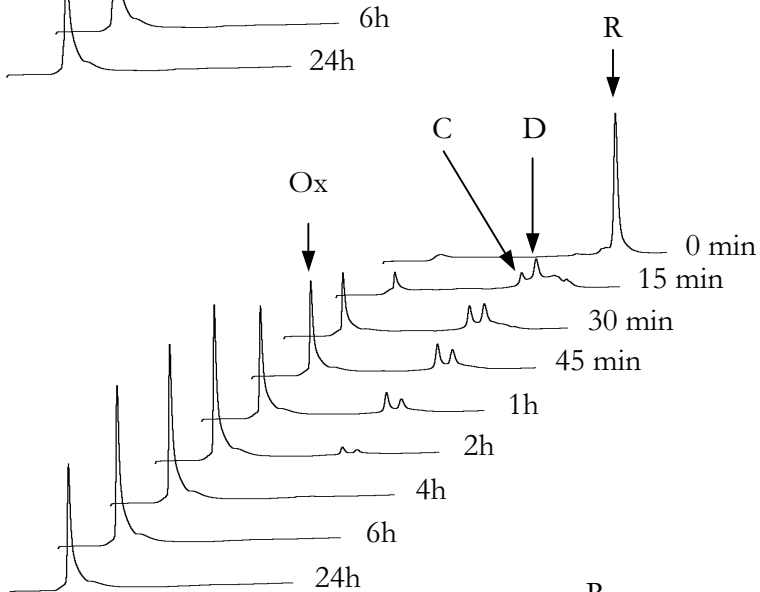
#### 1.1.2.3 Conclusion

La longueur de lk1 influe essentiellement sur la vitesse de conversion des intermédiaires C et D en forme native.

1



2



3

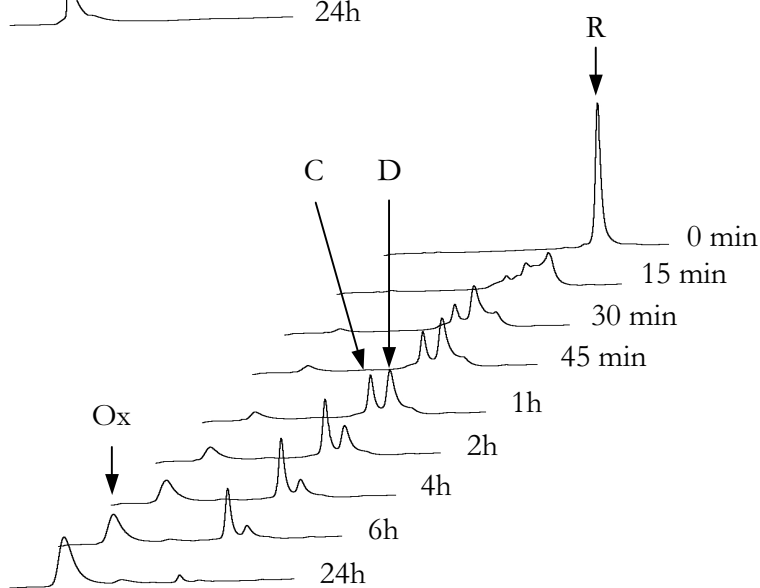
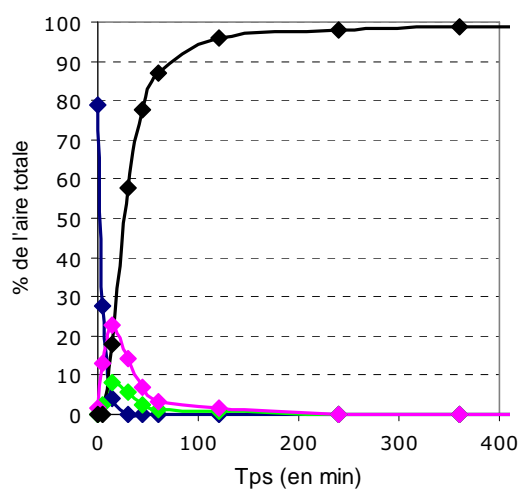


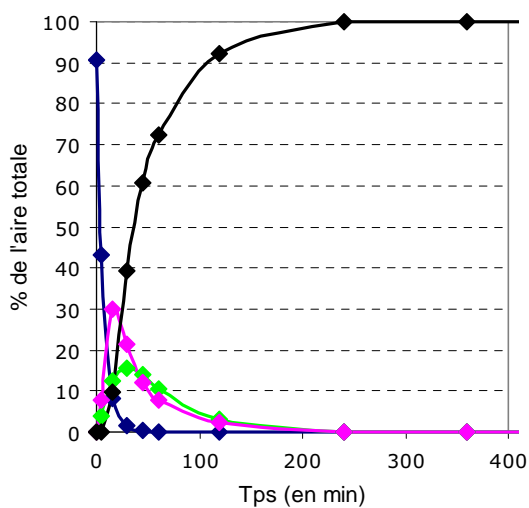
Figure 28 : Repliement oxydant de Tox60 (1), Tox61 (2) et Tox62 (3) aux temps longs, suivi en HPLC-RP.



1



2



3

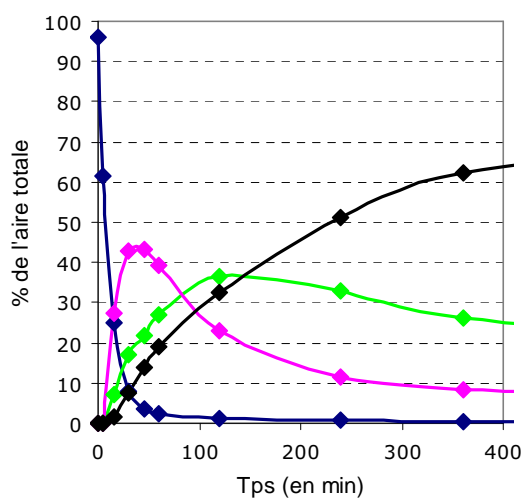


Figure 29 : Courbes de cinétique de repliement aux temps longs de Tox60 (1), Tox61 (2) et Tox62 (3). En bleu : la forme réduite ; en noir : la forme oxydée ; en vert : l'intermédiaire C et en magenta : l'intermédiaire D. Les courbes de cinétique ont été tracées en reliant les points expérimentaux.

## 1.2 La longueur de Ik1 détermine la productivité des intermédiaires à trois ponts

La suite de cette étude a pour but de décrire les phénomènes associés à la modification du comportement cinétique lors de l'étape finale du repliement. Le variant Tox60 produisant très peu d'intermédiaires à 3 ponts, et sachant qu'il nous est indispensable d'en obtenir des quantités importantes pour les caractériser, il a été décidé d'exclure ce variant du reste de l'étude. L'objectif sera donc de comparer le mécanisme de repliement entre la toxine  $\alpha$  naturelle (Tox61) et le variant possédant une insertion dans Ik1 (Tox62)

### 1.2.1 *Les intermédiaires C et D de Tox61 sont de même nature que ceux de Tox62*

Dans cette partie l'objectif est de caractériser les intermédiaires C et D de Tox61.

Les intermédiaires C et D se formant lors du repliement de Tox61 ont été bloqués à l'iodoacétamide puis purifiés. En collaboration avec Jean-Pierre LE CAER de l'Institut de Chimie des Substances Naturelles (ICSN, CNRS Gif-sur-Yvette) ces intermédiaires ont été étudiés en spectrométrie de masse.

#### 1.2.1.1 Détermination du nombre de ponts :

Les intermédiaires C et D de Tox61 bloqués ont été analysés en spectrométrie de masse. Dans les deux cas la masse correspond à celle de la forme native augmentée de 114 Da. Sachant que l'introduction d'un carboxylamide sur une cystéine libre provoque une augmentation de masse de 57 Dalton, ces intermédiaires ne possédaient que deux cystéines libres et représentaient donc des formes à trois ponts.

#### 1.2.1.2 Détermination des cystéines libres :

Ces intermédiaires ont été digérés à la trypsine puis réduits. Après dessalage sur micro colonne, les mélanges peptidiques ont été analysés en masse.

Résidus	Séquence	Masse monoisotopique
26-26	K	146,1
59-61	CNN	349,1
30-32	DHR	426,2
27-29	VWR	459,3
26-29	KVWR	587,4
33-38	GTIIER	687,4
27-32	VWRDHR	867,4
51-58	LNCCTTDK	896,4
26-32	KVWRDHR	995,5
30-38	DHRGTIIER	1095,6
16-25	TCPGETNCYK	1114,4
39-50	GCGCPTVKPGIK	1158,6
51-61	LNCCTTDKCNN	1227,5
16-26	TCPGETNCYKK	1242,5
27-38	VWRDHRGTIIER	1536,8
26-38	KVWRDHRGTIIER	1664,9
16-29	TCPGETNCYKKVWR	1683,8
1-15	LECHNQSSQPPTTK	1696,8
33-50	GTIIERGCGCPTVKPGIK	1828,0
30-50	DHRGTIIERGCGCPTVKPGIK	2236,2
39-61	GCGCPTVKPGIKLNCCTTDKCNN	2368,0
16-38	TCPGETNCYKKVWRDHRGTIIER	2761,4
26-50	KVWRDHRGTIIERGCGCPTVKPGIK	2805,5
1-26	LECHNQSSQPPTTKTCPGETNCYKK	2921,3
33-61	GTIIERGCGCPTVKPGIKLNCCTTDKCNN	3037,4
30-58	DHRGTIIERGCGCPTVKPGIKLNCCTTDK	3114,5
1-29	LECHNQSSQPPTTKTCPGETNCYKKVWR	3362,6
30-61	DHRGTIIERGCGCPTVKPGIKLNCCTTDKCNN	3445,6
27-58	VWRDHRGTIIERGCGCPTVKPGIKLNCCTTDK	3555,8
26-58	KVWRDHRGTIIERGCGCPTVKPGIKLNCCTTDK	3683,9
1-32	LECHNQSSQPPTTKTCPGETNCYKKVWRDHR	3770,7
27-61	VWRDHRGTIIERGCGCPTVKPGIKLNCCTTDKCNN	3886,9
16-50	TCPGETNCYKKVWRDHRGTIIERGCGCPTVKPGIK	3901,9
1-38	LECHNQSSQPPTTKTCPGETNCYKKVWRDHRGTIIER	4440,1

Tableau 7 : Liste des peptides susceptibles de se former lors de la digestion trypsique de Tox61

Le Tableau 8 présente les peptides observés lors de l'analyse de masse des digestions trypsiques. Seuls les peptides permettant de différencier l'intermédiaire C de l'intermédiaire D sont présentés.

Observation n°	Peptides théoriques dans la forme native		Peptides observés dans l'intermédiaire C		Peptides observés dans l'intermédiaire D	
	Résidus	Masse mono- isotopique	Masse observée	Nombre de CAM	Masse observée	Nombre de CAM
1	1-29	3362,6	3363,4	0	3420,5	1
2	1-32	3770,7	3771,5	0	3828,7	1
3	33-61	3037,4	3152,3	2	3095,4	1
4	39-61	2368,0	2482,1	2	2425	1
5	39-50	1158,6	1216,6	1	1216,6	1

Tableau 8 : Peptides observés dans les digestions trypsiques réduites et nombre de molécules de carboxyamidométhyl incorporées.

1



2



Figure 30 : Récapitulatif des informations obtenues par l'analyse par spectrométrie de masse des digestions réduites de l'intermédiaire C (1) et de l'intermédiaire D (2) de Tox61. Les numéros des observations correspondent à ceux de la première colonne du Tableau 8.

La Figure 30 indique que dans l'intermédiaire C de Tox61, une des cystéines 40 ou 42 est libre (observation 5), l'autre cystéine libre est soit 53, 54 ou 60 (Observations 3 et 4). Concernant l'intermédiaire D de Tox61, une des cystéines 3, 17 ou 23 est libre (Observations 1 et 2) et l'autre cystéine libre est soit 40, soit 42 (Observation 5).

Pour tenter d'identifier les cystéines carboxylamidés dans ces peptides, certains peptides ont été séquencés par fragmentation. Pour l'intermédiaire C, cette analyse n'a donné aucun résultat. Pour l'intermédiaire D, les séquençages des peptides 16-26 et 39-50 montrent que les cystéines 17 et 40 sont alkylées. D'après ces résultats, l'intermédiaire D de Tox61 a deux cystéines libres et ces dernières sont les mêmes que dans l'intermédiaire D de Tox62.

Pour tenter de lever le doute concernant les cystéines libres dans l'intermédiaire C de Tox61, les spectres de dichroïsme circulaire de cet intermédiaire bloqué et du mutant Tox61pt3 (avec les cystéines 42 et 53 mutées en alanine) ont été comparés (Figure 31).

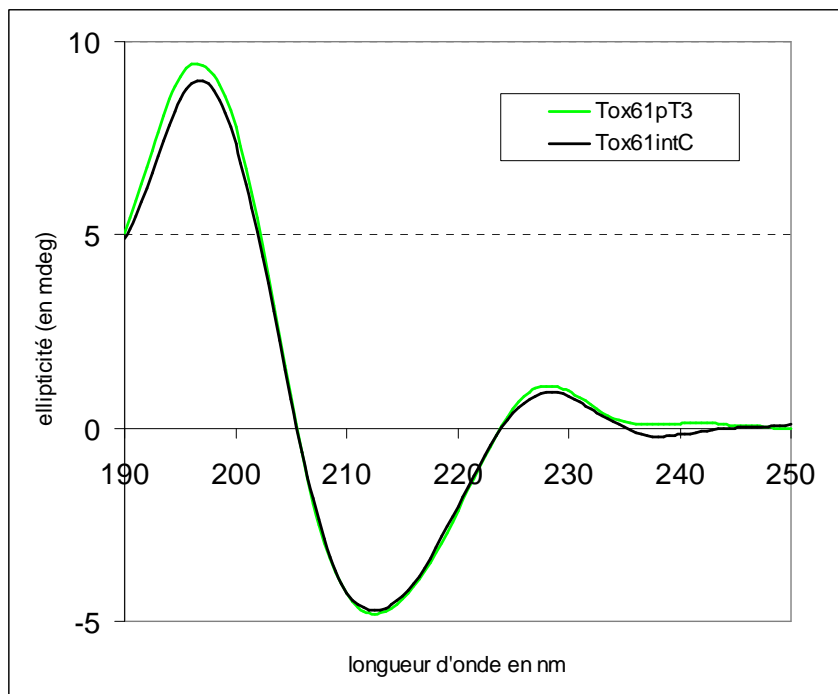


Figure 31 : Spectres de dichroïsme circulaire de Tox61pT3 et Tox61 intermédiaire C.

Les spectres de dichroïsme circulaire de Tox61pT3 et de l'intermédiaire C de Tox61 sont identiques. Les informations obtenues en analyse de masse associée à cette identité de contenu en structure secondaire constituent un faisceau de présomptions permettant de faire l'hypothèse que l'intermédiaire C de Tox61 est l'intermédiaire des[42-53] (voir la partie 3.3 de l'introduction pour la signification de cette notation).

#### 1.2.1.3 Conclusion

Les intermédiaires C et D de Tox61 sont donc des homologues des intermédiaires C et D de Tox62. L'ajout d'un résidu dans le lk1 n'a donc pas modifié la nature des formes intermédiaires s'accumulant lors du repliement oxydant de ces deux molécules.

### 1.2.2 Productivité des intermédiaires C et D par analyse de leur repliement

Maintenant que les intermédiaires de Tox61 ont été identifiés, le but est de déterminer lequel de ces intermédiaires à trois ponts est productif.

Le principe est d'isoler un intermédiaire et de le placer dans des conditions oxydantes pour qu'il se replie. En effet la méthode de blocage à l'acide n'est pas irréversible. Une fois bloqué un intermédiaire peut être purifié puis placé dans un tampon de renaturation. Il est alors possible de suivre son évolution comme cela a été effectué précédemment à partir de la protéine réduite. Dans ce cas, l'objectif étant de tester la productivité des intermédiaires, nous avons choisi d'utiliser un tampon de repliement contenant uniquement du glutathion oxydé. Dans ces conditions, il est possible de déterminer si un intermédiaire génère directement la forme native ou s'il est nécessaire qu'un réarrangement (nécessairement intramoléculaire puisque le milieu de repliement ne contient pas de réducteur) ait lieu. Ruoppolo *et al.* avaient réalisé ce type d'expérience pour les intermédiaires C et D de Tox62 ((Ruoppolo et al., 2001)). Leurs résultats montraient que C formait directement la forme native alors que l'intermédiaire D devait nécessairement se transformer au préalable en l'intermédiaire C pour pouvoir ensuite produire la forme native. Ces auteurs en avaient conclu que lors du repliement de Tox62, seul l'intermédiaire C est productif.

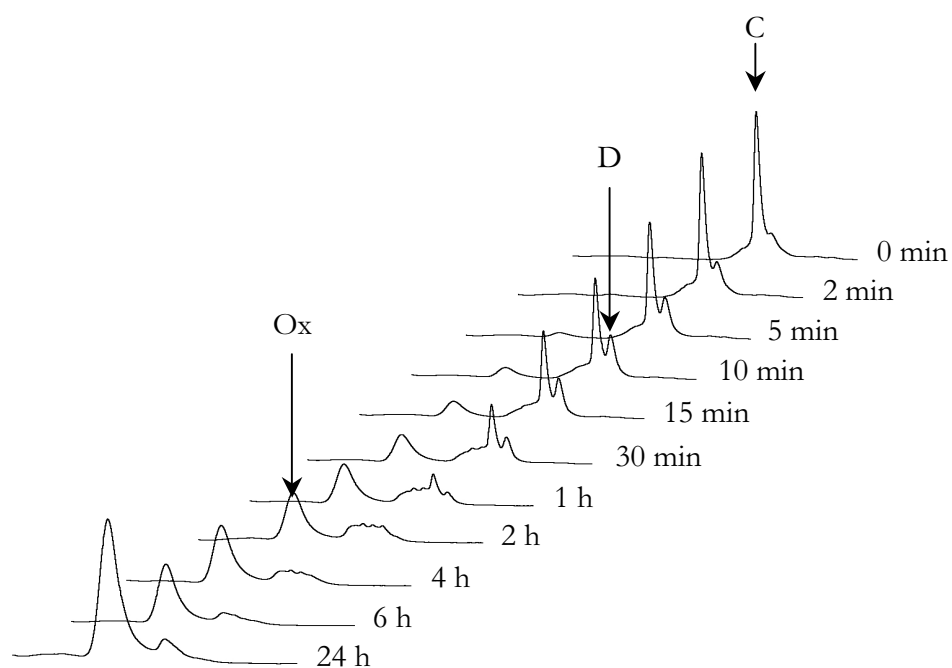
#### 1.2.2.1 Evolution des profils chromatographiques en phase inverse des mélanges de repliement effectués à partir des intermédiaires purifiés

Les chromatogrammes présentés en Figure 32 décrivent les repliements des intermédiaires C et D de Tox61 et de Tox62 isolés.

Dans le cas de Tox61, que l'on parte de l'intermédiaire C ou de l'intermédiaire D, dans les deux cas l'autre intermédiaire apparaît rapidement. Lors du repliement à partir de 100% de forme D, la forme native apparaît dès les premiers temps de repliement ce qui n'est pas le cas lorsqu'on part de 100% de forme C. De plus dans les deux cas, au temps 30 minutes l'intermédiaire le plus abondant est l'intermédiaire C. Ces observations suggèrent que l'intermédiaire C est peu productif. Des pics ayant des temps de rétention inférieurs à celui de l'intermédiaire C apparaissent dans les deux cas et s'accumulent. Ces molécules seront classées comme formes indéterminées. Les mêmes expériences ont été réalisées dans un tampon contenant le couple oxydo-réducteur glutathion oxydé/glutathion réduit (chromatogrammes non montrés). Dans ces conditions, les formes indéterminées sont nettement moins abondantes.

Dans le cas de Tox62, on observe que l'intermédiaire D produit plus facilement l'intermédiaire C que l'inverse. D'autre part la forme D semble devoir se transformer en la forme C pour que la forme N puisse apparaître (Figure 33 1 et 2). Puisque ces expériences sont menées en absence de réducteur, l'échange se fait par « reshuffling » (voir partie 3.1.1 de l'introduction). Contrairement à ce qui avait été observé par Ruoppolo *et al.*, il existe un échange entre les formes C et D.

1



2

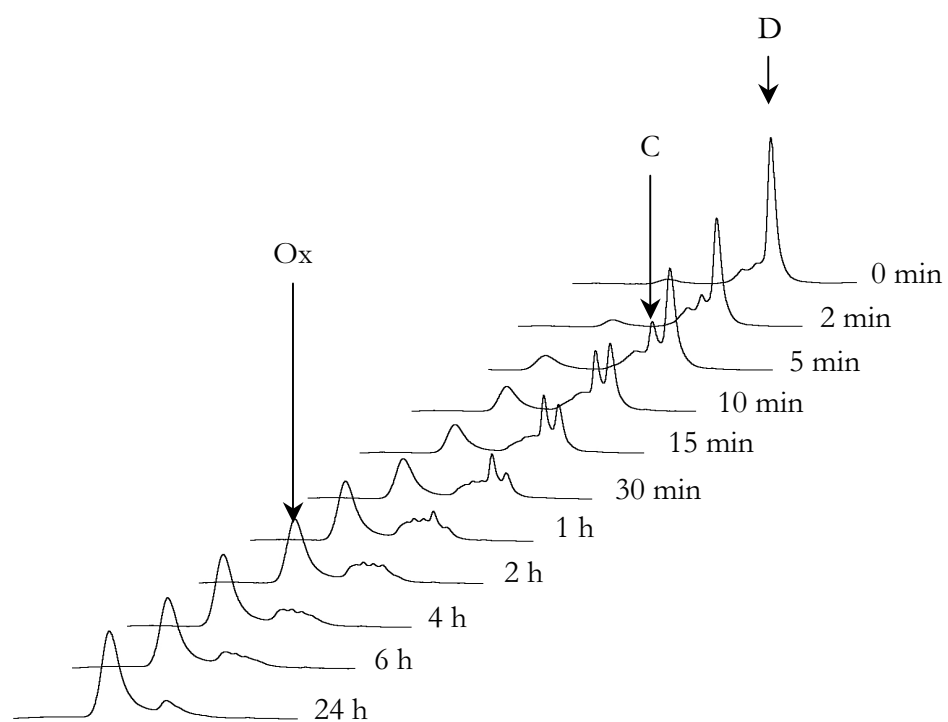
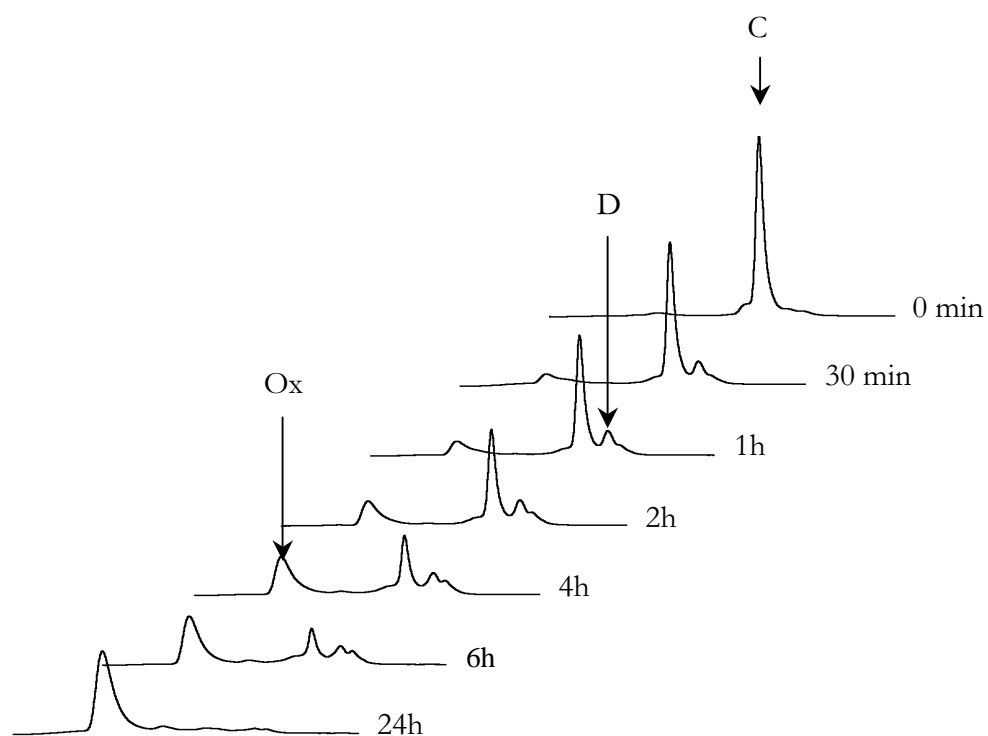


Figure 32 : Cascade de profils chromatographiques de repliement à partir de 100% de forme C (1) ou de forme D (2) de Tox61 en condition uniquement oxydante et suivie par HPLC phase inverse.

1



2

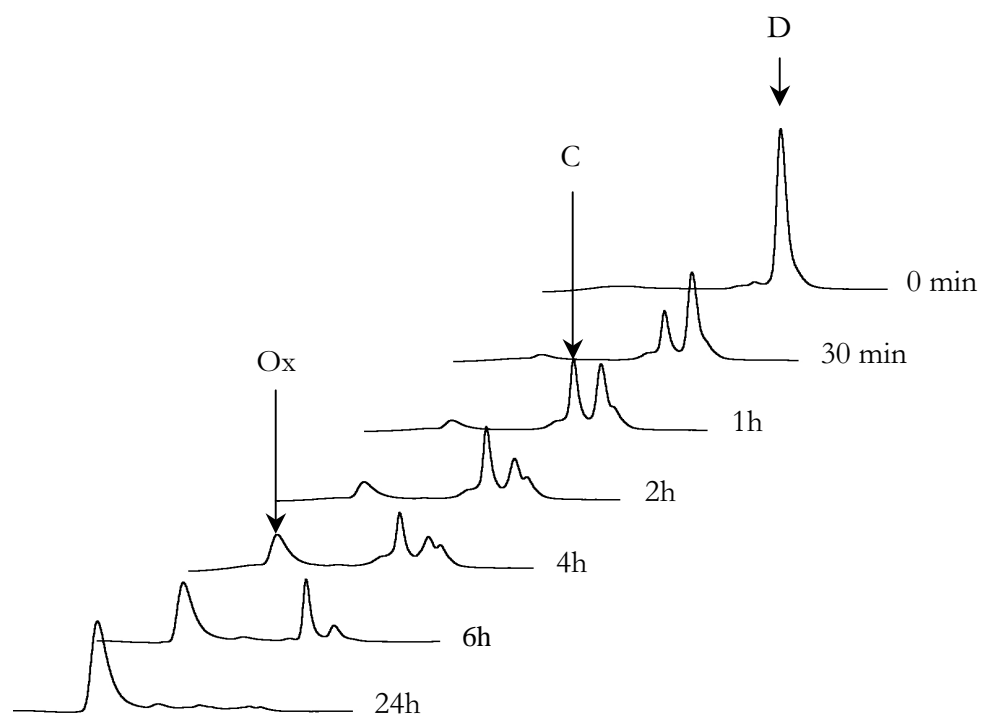


Figure 33: Cascade de profils chromatographiques de repliement à partir de 100% de forme C (1) ou de forme D (2) de Tox62 en condition uniquement oxydante et suivie par HPLC phase inverse.



### 1.2.2.2 Intégration et courbes cinétiques

L'intégration des chromatogrammes précédents a permis d'obtenir les points expérimentaux présentés en Figure 35 1, 2, 3, et 4.

Dans le cas de Tox62 (Figure 35 3 et 4), la forme C produit plus rapidement la forme native que la forme D. Le temps de formation de 50% de forme oxydée est de 320 minutes pour l'intermédiaire C et 500 minutes pour l'intermédiaire D. Il semble donc que l'intermédiaire C est plus productif que l'intermédiaire D.

Dans le cas de Tox61 (Figure 35 1 et 2), lorsqu'on réalise le repliement à partir de 100% d'intermédiaire C ou de 100% de D, le temps d'apparition de 50% de forme native est respectivement de 50 minutes et de 35 minutes. La situation semble donc inversée par rapport à ce qui a été observé pour Tox62.

### 1.2.2.3 Modélisation du mécanisme

Les résultats du repliement en condition uniquement oxydante à partir de 100% de forme C ou D de Tox61 ou de Tox62 ont été modélisés (Figure 35). Ces expériences ont montré que les formes C et D sont en échange et qu'il existe d'autres formes non déterminées dont il faut tenir compte (espèce X). La molécule de départ étant un intermédiaire à 3 ponts (intermédiaire C ou D) et sachant que les expériences de repliement ont été menées en l'absence de réducteur, ces formes indéterminées ont nécessairement un nombre de ponts supérieur ou égal à trois. De plus, il a été observé que ces espèces indéterminées s'accumulent (Figure 32), elles sont considérées comme non productives.

D'après ces données, on peut établir le modèle représenté dans la Figure 34 et le système d'équations différentielles qui en résulte.

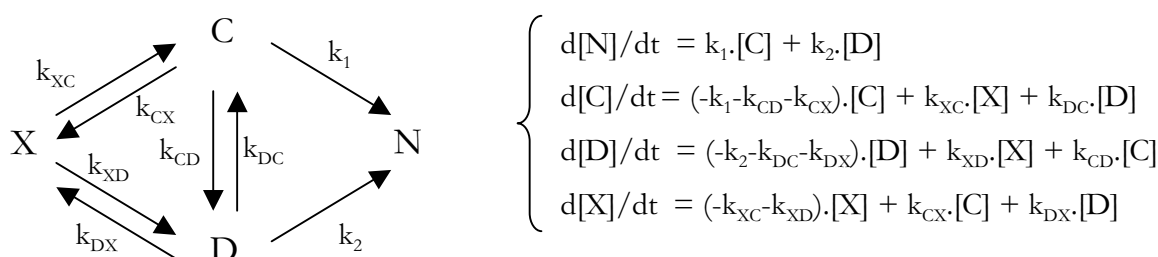


Figure 34 : A gauche, le mécanisme du repliement à partir des intermédiaires à trois ponts et en condition uniquement oxydante et à droite, les équations différentielles associées à ce mécanisme.

Nous avons utilisé le logiciel Madonna Berkeley pour résoudre ce système d'équations différentielles et trouver les valeurs des constantes de vitesse permettant d'expliquer au mieux les points expérimentaux obtenus (Figure 35 et Tableau 9). L'algorithme de résolution numérique Runge Kutta a été utilisé ((Rothwarf and Scheraga, 1993)).

L'analyse de la Figure 35 montre que la simulation effectuée avec le modèle présenté en Figure 34 décrit correctement le comportement cinétique dans le cas de Tox61 comme dans le cas de Tox62 que l'on parte de 100% de l'intermédiaire C ou de 100% de l'intermédiaire D.

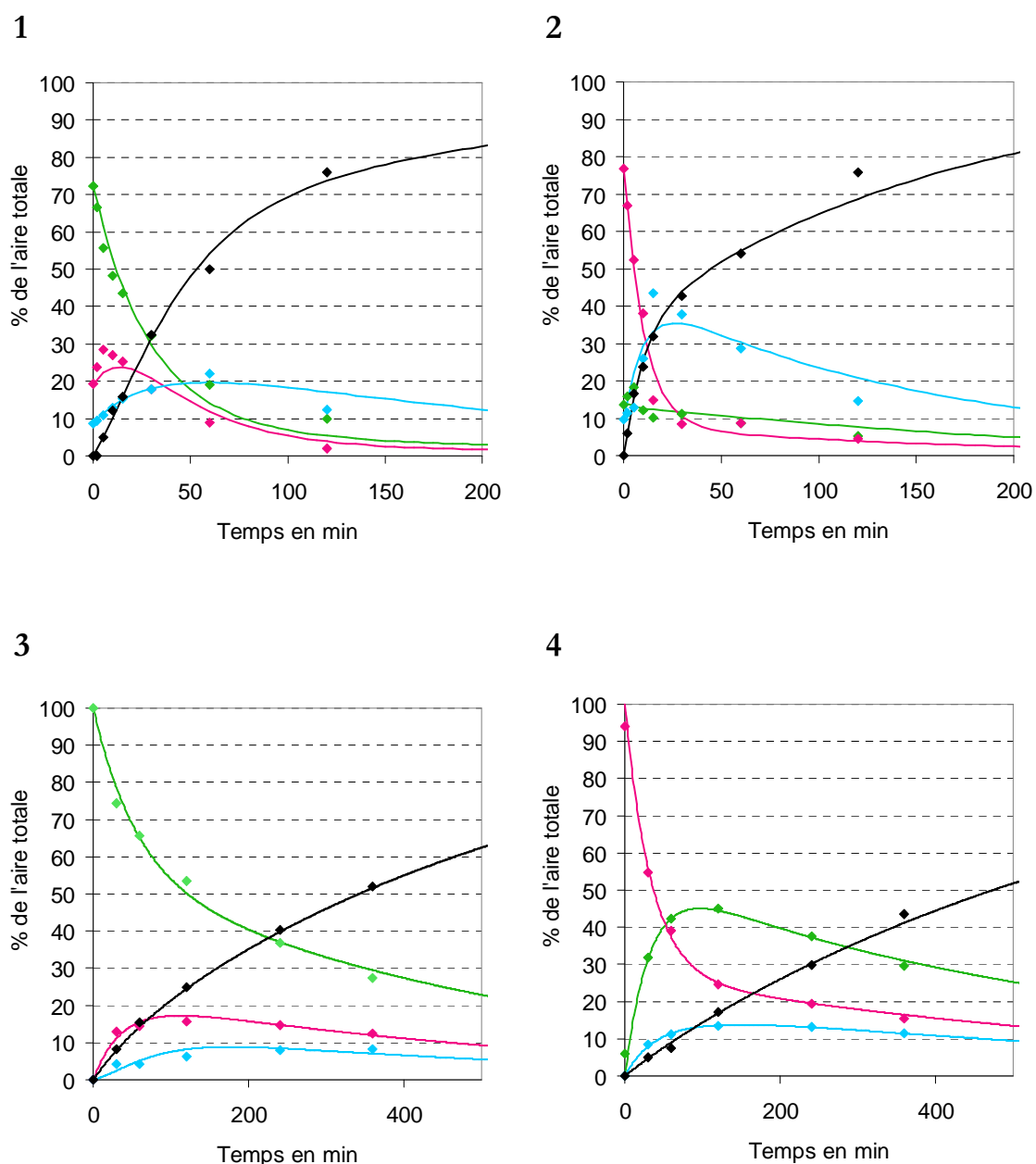


Figure 35 : Points expérimentaux et courbes de modélisation issus des cinétiques de repliement effectuées à partir de 100% (ou 80% pour Tox61) des intermédiaires C et D. 1 : Tox61 à partir de 80% de l'intermédiaire C ; 2 : Tox61 à partir de 80% de l'intermédiaire D ; 3 : Tox62 à partir de 100% de l'intermédiaire C ; 4 : Tox62 à partir de 100% de l'intermédiaire D. En noir : la forme oxydée ; en vert : l'intermédiaire C ; en magenta : l'intermédiaire D ; en cyan : formes indéterminées.

A partir des optimisations obtenues avec Berkeley Madonna, il est possible de calculer les constantes de vitesse du modèle (Tableau 9). Il existe une incertitude sur les mesures basées sur l'intégration des pics de chromatographie, en particulier dans l'estimation des formes peu abondantes. Cette incertitude peut expliquer l'hétérogénéité constatée dans la détermination des constantes selon que l'on part de 100% de l'un ou l'autre intermédiaire. En conséquence, pour une réaction donnée, on considérera comme fiable la valeur de constante déterminée à partir de l'expérience dans laquelle l'espèce de départ correspond à la constante étudiée, soit 100% de C pour k1 et 100% de D pour k2 (valeurs surlignées en jaune dans le Tableau 9)

	Tox61		Tox62	
	100%C	100%D	100%C	100%D
<b>k1</b>	<b>5,6.10<sup>-9</sup></b>	<b>7,2.10<sup>-7</sup></b>	<b>3,0.10<sup>-3</sup></b>	<b>1,8.10<sup>-3</sup></b>
<b>k2</b>	<b>4,7.10<sup>-2</sup></b>	<b>4,9.10<sup>-2</sup></b>	<b>1,2.10<sup>-9</sup></b>	<b>1,7.10<sup>-3</sup></b>
kCD	2,6.10 <sup>-2</sup>	5,4.10 <sup>-3</sup>	5,6.10 <sup>-3</sup>	7,5.10 <sup>-3</sup>
kDC	4,0.10 <sup>-3</sup>	1,3.10 <sup>-5</sup>	1,1.10 <sup>-2</sup>	1,7.10 <sup>-2</sup>
kDX	2,8.10 <sup>-4</sup>	4,2.10 <sup>-2</sup>	1,0.10 <sup>-2</sup>	3,9.10 <sup>-3</sup>
kXD	5,7.10 <sup>-9</sup>	1,4.10 <sup>-2</sup>	9,7.10 <sup>-3</sup>	9,5.10 <sup>-3</sup>
kCX	8,1.10 <sup>-3</sup>	1,1.10 <sup>-2</sup>	4,5.10 <sup>-4</sup>	9,9.10 <sup>-4</sup>
kXC	6,2.10 <sup>-3</sup>	4,2.10 <sup>-3</sup>	1,0.10 <sup>-2</sup>	4,4.10 <sup>-8</sup>

Tableau 9 : Constantes de vitesse obtenues par simulation numérique grâce au modèle de la Figure 34 (les valeurs sont exprimées en min<sup>-1</sup>)

La modélisation du repliement à partir des intermédiaires de Tox61 montre que la constante de vitesse associée à la conversion de la forme C en forme native est quasi-nulle. A l'inverse, la constante associée à l'intermédiaire D de Tox61 est très élevée, cette voie est donc la voie productive du repliement de Tox61.

Dans le cas de Tox62, les vitesses de conversion de l'intermédiaire C et de l'intermédiaire D en forme native sont plus de dix fois inférieures à celle de la voie productive de Tox61 ; cette différence explique la différence de cinétique globale entre Tox61 et Tox62. La voie privilégiée de repliement de Tox62 est celle empruntant l'intermédiaire C puisque sa vitesse de conversion en forme native est deux fois plus rapide que celle de l'intermédiaire D.

#### 1.2.2.4 Conclusion

En comparant les constantes d'un mutant à l'autre, il apparaît que l'ajout d'un résidu dans lk1 implique deux modifications dans le mécanisme de repliement:

- Une voie sans issues dans Tox61 (intermédiaire C) a été rendue productive dans Tox62.
- La voie productive du repliement de Tox61 (par l'intermédiaire D) est moins efficace dans le cas de Tox62.

La suite de cette étude a pour but d'obtenir une explication structurale à la première modification.

## 1.3 Interprétation structurale des différences de mécanisme de repliement

### 1.3.1 Les structures cristallographiques de Tox61 et Tox62

La structure cristallographique de Tox61 a été publiée en 2003 ((Gilquin et al., 2003)). La structure cristallographique de Tox62 à une résolution de 2,1 Å a été résolue par Enrico Stura et Renée Ménez dans le cadre de cette étude.

L'alignement structural effectué avec le logiciel PYMOL (Figure 36) fournit un RMSD de 1 Å sur 459 atomes. L'analyse de ces structures montre qu'elles sont pour la majorité des résidus, parfaitement identiques. Seule la zone agrandie présentée sur la Figure 36, correspondant au lk1, diffère significativement. C'est dans cette boucle qu'un résidu sérine a été ajouté dans le mutant Tox62 (en rouge). Cette mutation n'a pas eu d'effet longue distance sur la structure de la toxine, seule une adaptation locale de la présence d'un résidu supplémentaire dans cette boucle explique les légères différences structurales observées.

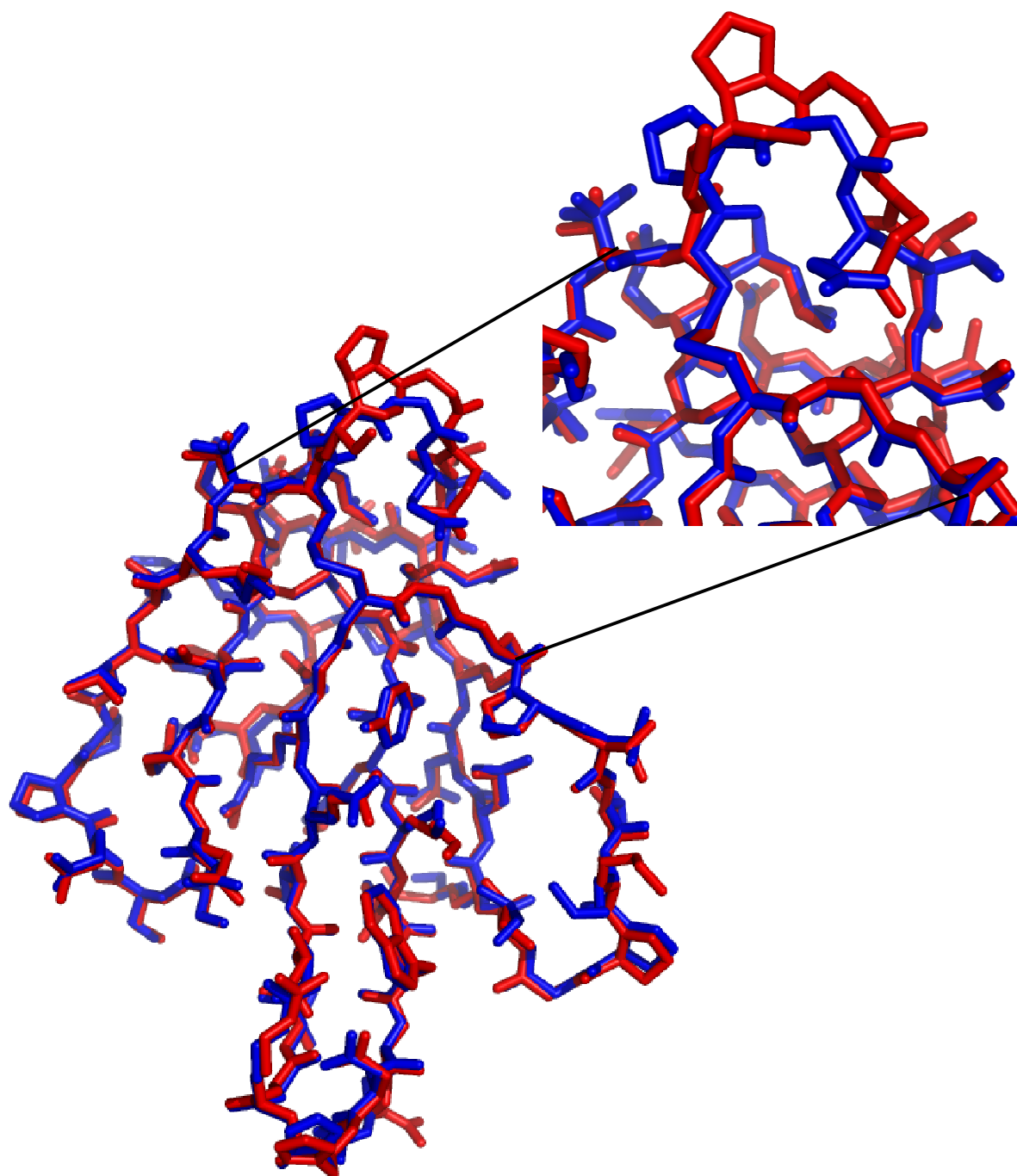


Figure 36 : Comparaison des structures cristallographiques de Tox61 (en bleu) et Tox62 (en rouge). La figure de droite est un agrandissement du lk1. L'alignement structural a été produit par le programme PYMOL (<http://pymol.sourceforge.net/>).

### 1.3.2 Les variants Tox61pT3 et Tox62pT3

Pour des raisons pratiques, il est difficile de produire en quantité importante les intermédiaires C de Tox61 et de Tox62. Nous avons donc opté pour l'utilisation de variants pour lesquels les cystéines correspondant aux ponts manquants sont substituées par des alanines. Les variants Tox61C42A/C53A et Tox62C43A/54A seront par la suite désignés respectivement par les termes de Tox61pT3 et Tox62pT3. Ces protéines ont été considérées comme des analogues structuraux des intermédiaires C de Tox61 et Tox62 (voir la partie 3.4 de l'introduction pour une discussion sur l'utilisation de ces mutants pour analyser la structure des intermédiaires de repliement des protéines riches en ponts disulfure).

### 1.3.3 Le mutant Tox61pT3 dispose d'une stabilité thermique plus faible que celle de Tox62pT3

#### 1.3.3.1 Structure secondaire de Tox61pT3 et Tox62pT3

La structure secondaire de ces variants a été étudiée par dichroïsme circulaire.

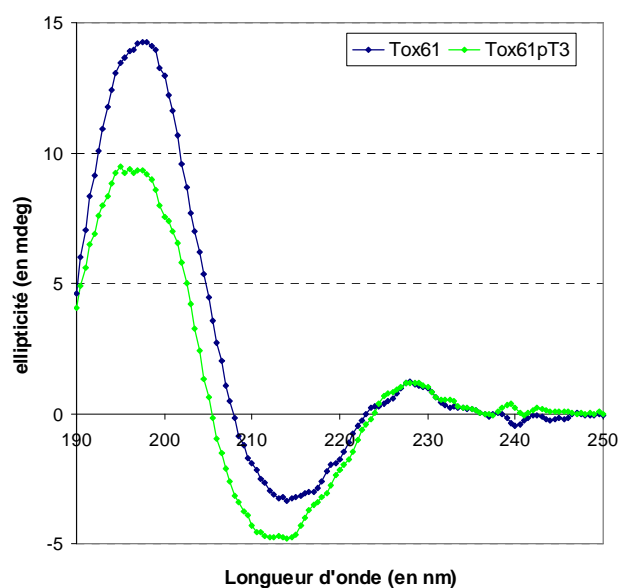
Les spectres de dichroïsme circulaire de Tox61pT3 et Tox62pT3 montrent que ces protéines sont structurées en feuillet  $\beta$  (bandes à 198 nm et 215nm : Figure 37). Dans l'ensemble, la comparaison des spectres de ces variants avec ceux des protéines natives leur correspondant indique que la structure secondaire est conservée lorsque les cystéines du pont 3 sont mutées en alanine.

Cependant, il existe une diminution de l'intensité de la bande à 198 nm signifiant que dans les mutants de pont 3 la quantité moyenne de structure  $\beta$  est plus faible. Cette différence est plus prononcée dans le cas de Tox62pT3. Si l'on compare les spectres de dichroïsme circulaire de Tox61 et de Tox62, on observe qu'ils sont similaires excepté au niveau de la bande à 228 nm qui est trois fois plus importante dans le cas de Tox62. L'existence de cette bande a été attribuée à la présence de ponts disulfures dont les angles dièdres au niveau des ponts sont proches de 90°. ((Hider et al., 1988), (Rauk, 1984)). Il est probable que les angles dièdres des cystines de Tox61 soient légèrement différents de ceux de Tox62 ce qui aboutit à cette modification de l'intensité de la bande à 228nm.

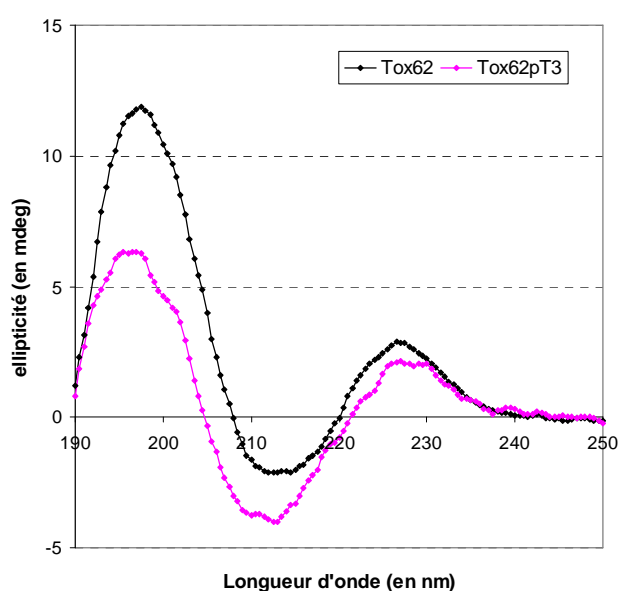
#### 1.3.3.2 Dénaturation thermique suivie en dichroïsme circulaire

Cette expérience décrit l'évolution de la structure secondaire (ici les feuillets  $\beta$ ) lors de la dénaturation thermique de la molécule. Les courbes montrent que les deux mutants atteignent un signal de protéine dénaturée équivalent, ce qui tend à prouver que les états dénaturés finaux sont similaires. La température de fusion ou  $T_m$  de ces protéines a été estimée graphiquement, elle est de 43°C pour Tox61pT3 et de 52°C pour Tox62pT3. Tox61pT3 est donc moins stable que Tox62pT3.

1



2



3

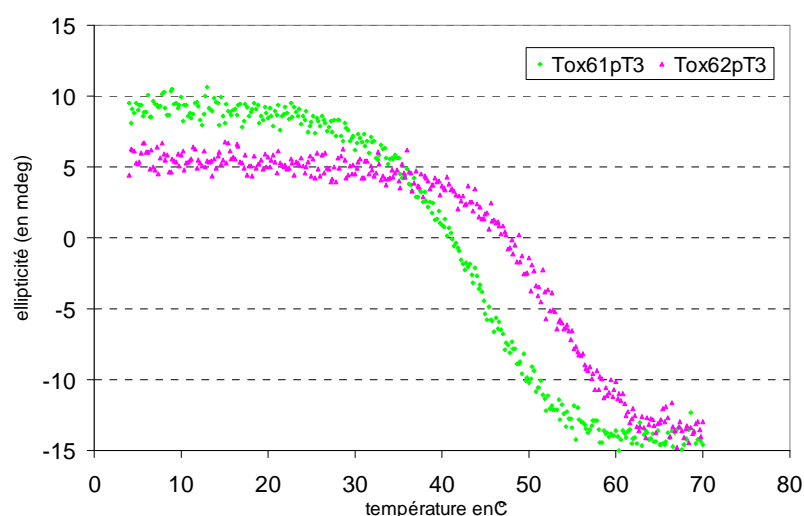


Figure 37 : Spectres de dichroïsme circulaire de Tox61, Tox62 et de leurs mutants de pont 3 respectifs (1) et (2). Suivi de la dénaturation thermique des mutants Tox61pT3 et Tox62pT3 en dichroïsme circulaire à 198nm (3).

### 1.3.4 Structure tridimensionnelle étudiée en RMN des variants Tox61pT3 et Tox62pT3

#### 1.3.4.1 Attribution des déplacements chimiques des atomes de Tox61, Tox62, Tox61pT3, Tox62pT3

Les expériences suivantes ont été effectuées sur l'ensemble des mutants :  $^1\text{H}$   $^{15}\text{N}$  HSQC,  $^1\text{H}$   $^1\text{H}$  COSY 2D,  $^1\text{H}$   $^1\text{H}$  TOCSY 2D,  $^1\text{H}$   $^1\text{H}$   $^{15}\text{N}$  TOCSYHSQC 3D,  $^1\text{H}$   $^1\text{H}$  NOESY 2D,  $^1\text{H}$   $^1\text{H}$   $^{15}\text{N}$  NOESYHSQC 3D. A partir de ces expériences, et en utilisant l'attribution effectuée par Sophie ZINN-JUSTIN sur la toxine  $\alpha$  (appelée ici Tox61) dans des conditions de pH différentes ((Zinn-Justin et al., 1992)), les déplacements chimiques de chacun des atomes d'azote 15 et des protons de ces variants ont pu être déterminés (Figure 37). Les listes de déplacements chimiques pour ces quatre protéines sont fournies en annexe 4.

L'analyse des spectres  $^1\text{H}$   $^{15}\text{N}$  HSQC de Tox61, Tox61pT3, Tox62, Tox62pT3 montre que de nombreux pics sont superposés entre les toxines natives et leurs mutants de pont 3 (Figure 38). A ce niveau d'analyse, cela constitue déjà une indication que l'ensemble de la structure est conservée. Une deuxième information peut être déduite: le spectre de Tox61pT3 (en vert) contient deux fois plus de pics que celui de Tox61 (en bleu). Cela suggère que Tox61pT3 est composé de deux populations dont la structure diffère de manière significative. La moitié des pics sont superposables (ou proches de l'être) avec ceux du spectre de Tox61. Cette forme est baptisée Tox61pT3 forme A et sera analysée dans cette partie. La deuxième forme dont les déplacements chimiques sont plus éloignés de ceux de Tox61 est appelée Tox61pT3 forme B et étudiée dans la partie 1.3.5 de cette partie résultats.

#### 1.3.4.2 Différences de déplacement chimique

Les histogrammes présentés sur les Figures 39 et 40 mettent en évidence les résidus les plus affectés par la substitution des cystéines du pont 3 par des alanines.

D'une manière générale les déplacements chimiques sont peu modifiés par la mutation des cystéines du pont 3.



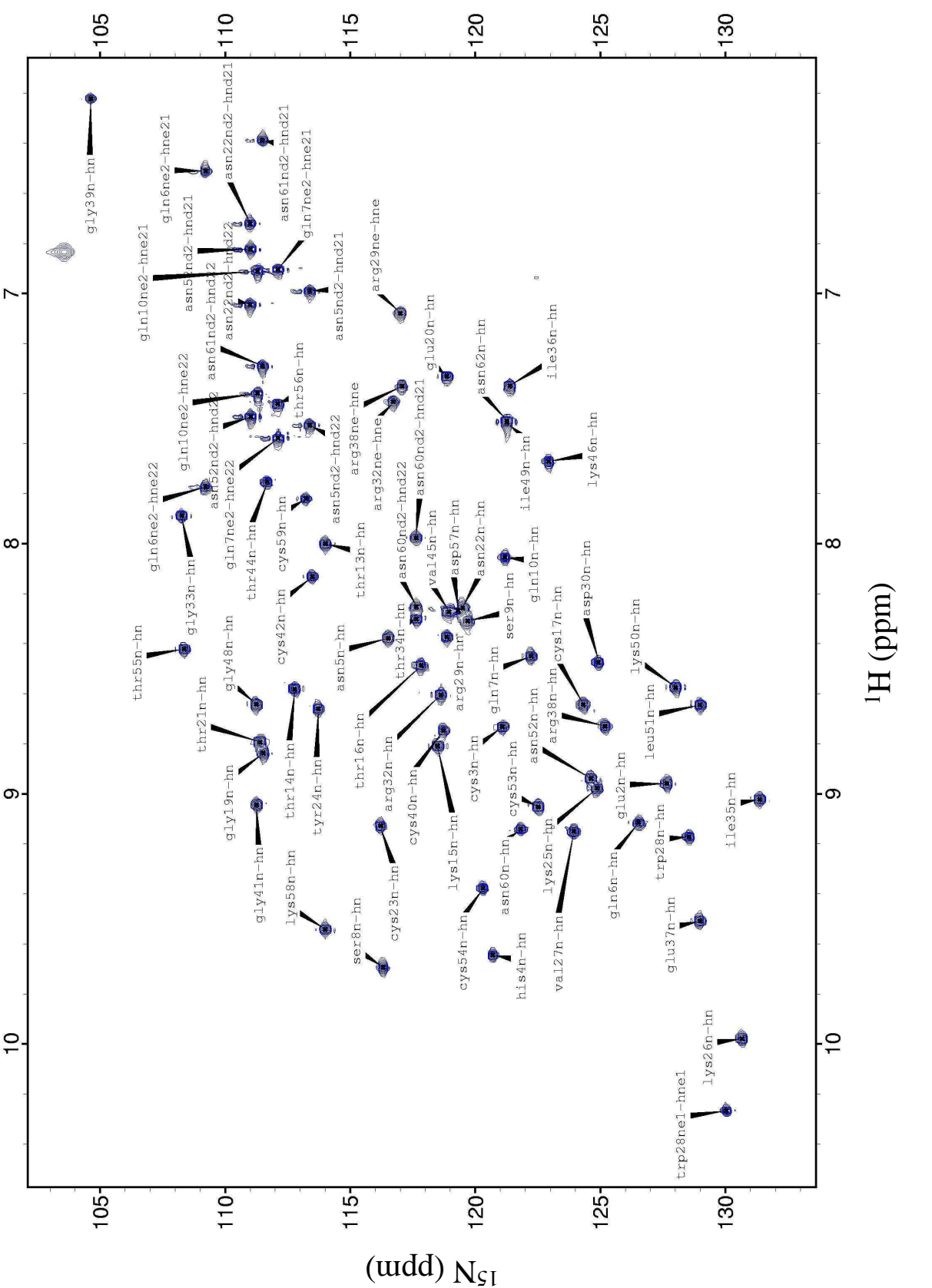


Figure 37 : Attribution du spectre  $^1\text{H}$   $^{15}\text{N}$  HSQC de Tox61 à pH5.

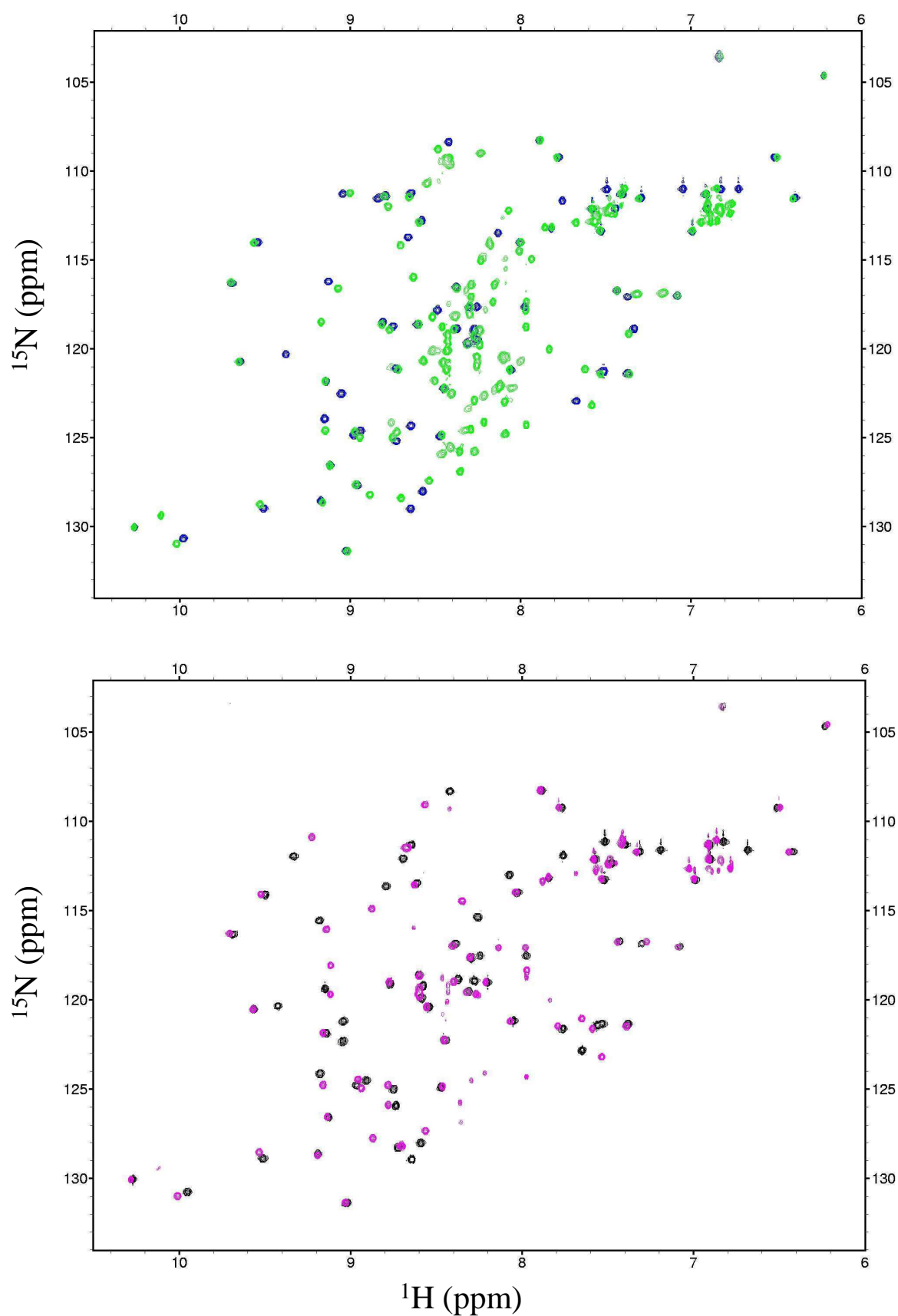


Figure 38 : Superposition de spectres  $^1\text{H}$   $^{15}\text{N}$  HSQC. 1 : Tox61 (en bleu) et Tox61pT3(en vert). 2 : Tox62 (en noir) et Tox62pT3 (en magenta).

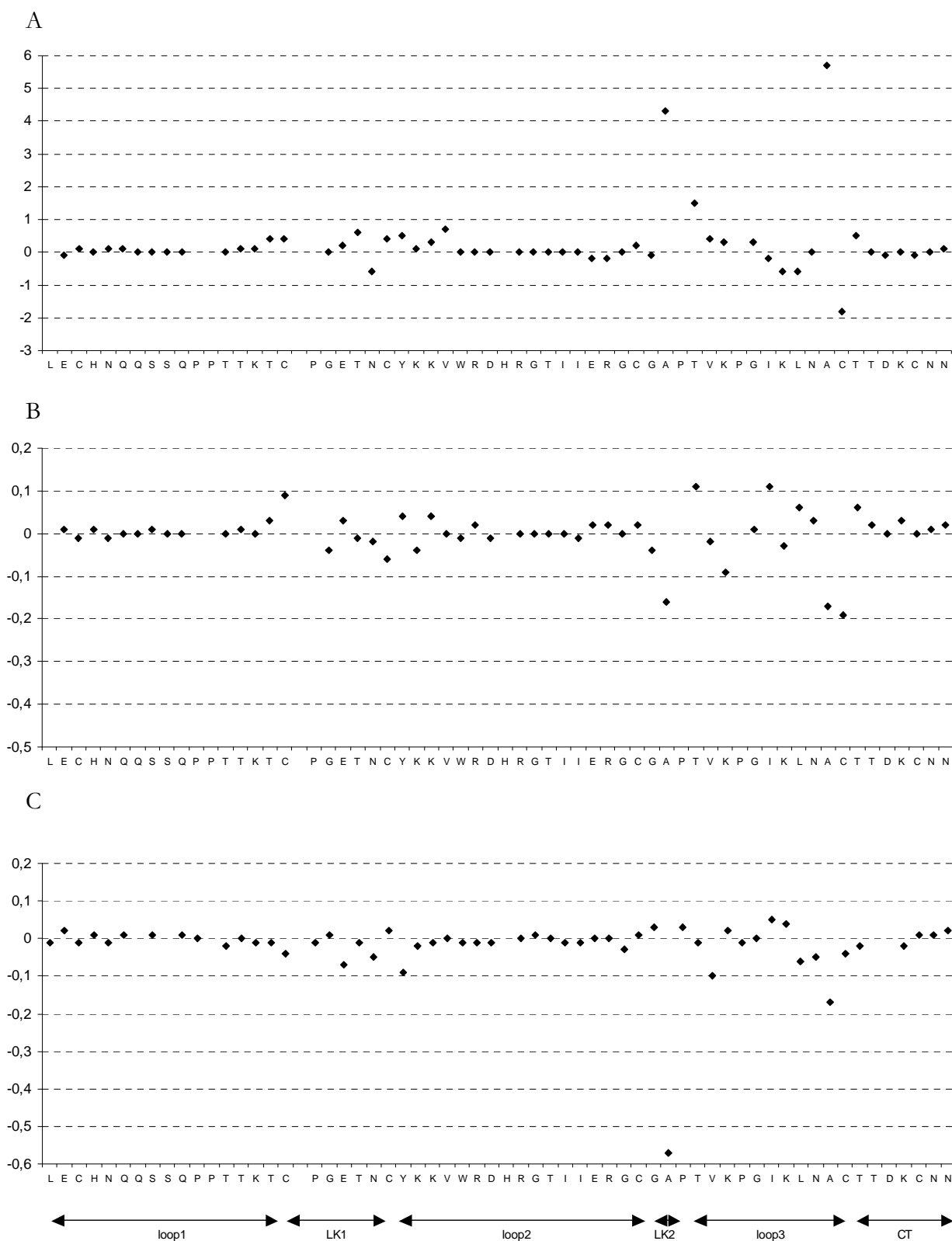


Figure 39 : En abscisse, les acides aminés de la séquence de Tox61pT3. En ordonnée, les différences de déplacement chimique (en ppm) entre Tox61 et Tox61 pT3 (forme A) correspondantes. A : Azote 15 du squelette peptidique ; B : Proton amide du squelette peptidique ; C : Proton lié au carbone  $\alpha$ .

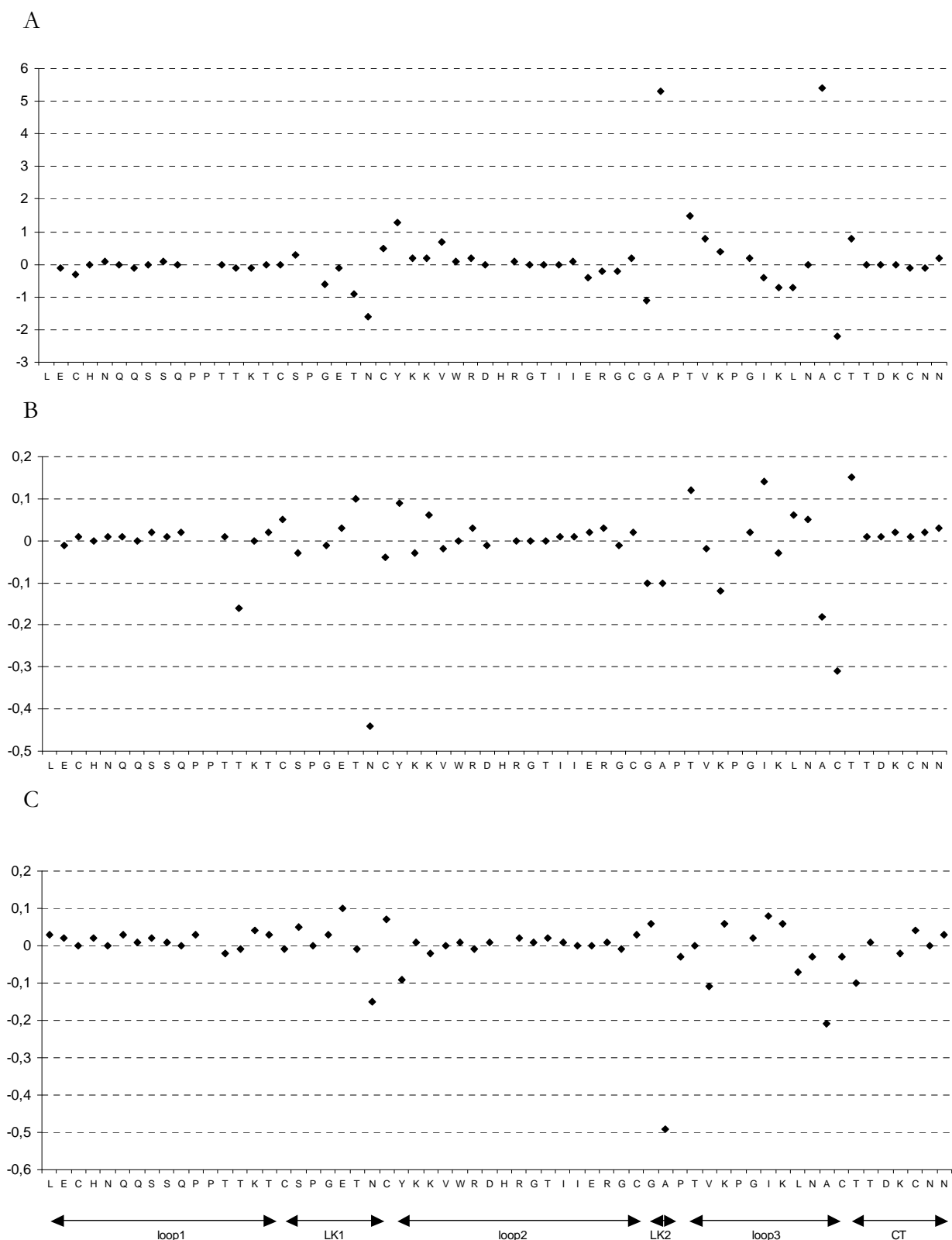


Figure 40 : En abscisse, les acides aminés de la séquence de Tox62pT3. En ordonnée, les différences de déplacement chimique (en ppm) entre Tox62 et Tox62pT3 (forme A) correspondantes. A : Azote 15 du squelette peptidique ; B : Proton amide du squelette peptidique ; C : Proton lié au carbone  $\alpha$ .

Même s'il est délicat d'interpréter directement en terme de structure la valeur d'un déplacement chimique, ((Pardi et al., 1983)), le fait que les atomes de deux molécules partagent des valeurs de déplacement chimique identiques indique que leurs environnements chimiques et donc leurs positions dans la structure sont similaires. Aussi bien dans le contexte à 61 résidus (Figure 39 A à C) que dans celui à 62 résidus (Figure 40 A à C), la plupart des déplacements chimiques des mutants de pont sont identiques à ceux des protéines à quatre ponts. Les résidus les plus affectés sont les alanines qui correspondent aux mutations. Evidemment, en remplaçant une cystéine par une alanine, non seulement la possibilité qu'un pont se forme est éliminée mais l'environnement chimique des atomes du squelette peptidique de ces résidus est affecté par la modification de leur chaîne latérale. De plus les atomes situés à proximité de l'alanine dans la structure tridimensionnelle de la protéine peuvent voir leur environnement chimique modifié même s'il n'y a pas eu de modification de la structure.

Chaque pic observé sur le spectre  $^1\text{H}$   $^1\text{H}$  NOESY réalisé dans l'eau correspond à un couple de protons situés à moins de 5 Å l'un de l'autre. Afin de connaître l'origine des modifications de déplacement chimique dans les mutants pont 3, les environnements des atomes des résidus dont les déplacements chimiques du squelette ont été le plus modifiés par la mutation ont été comparés entre les protéines à quatre ponts (Tox61 et Tox62) et leurs mutants de pont 3 respectifs (Tox61pT3 et Tox62pT3). Pour le contexte à 61 résidus ont été étudiés : N22, C ou A42, C ou A53, T21, Y24, V27, T44, V45, K50, L51, C54, T55, E20 et dans le contexte à 62 résidus : N23, C ou A43, C ou A 54, T56, V46, E21, Y25, I50, L52, T45, K47, K51, G42, V28, T22, C55 et C24. L'annexe 5 présente les résultats de cette analyse.

Il apparaît que tous les résidus dont le déplacement chimique a subi une importante modification sont à proximité des même atomes dans la structure des protéines à quatre ponts et dans la structure du mutant de pont 3 correspondant. Dans le contexte Tox61 seule la leucine 51 a subi une modification qui ne peut être due à la substitution des cystéines du pont 3 en alanine : dans le cas de Tox61pT3 forme A, sa chaîne latérale ne semble plus orientée vers l'intérieur de la boucle 3. Dans le contexte Tox62, aucune modification significative de la position des résidus n'a pu être observée. La forme A de Tox61pT3 et Tox62pT3 sont donc tous deux structurés comme Tox61 et Tox62 respectivement (exceptée la chaîne latérale de la leucine 51 de Tox61pT3 forme A). Les structures cristallographiques décrites en partie 1.3.1, sont de bons modèles de la structure des mutants Tox61pT3 forme A et Tox62pT3. Les structures tridimensionnelles de Tox61 et Tox62 étant voisines (Figure 36), Tox61 pT3 forme A et Tox62pT3 ont également des structures tridimensionnelles proches.

#### 1.3.4.3 Conclusion

La conformation A de l'intermédiaire C de Tox61 et l'intermédiaire C de Tox62 sont structurés comme leurs formes natives respectives. L'unique différence structurale significative entre ces intermédiaires est l'existence d'une deuxième conformation dans l'intermédiaire C de Tox61.

### 1.3.5 Analyse structurale de la forme B de Tox61pT3

#### 1.3.5.1 Attribution du spectre $^1\text{H}$ $^{15}\text{N}$ HSQC de Tox61pT3 forme B

L'analyse du spectre  $^1\text{H}$   $^1\text{H}$  NOESY 2D dans l'eau montre qu'il existe des pics d'échange entre les protons amides du squelette peptidique de Tox61pT3 qui ont été attribués à la forme A et des atomes non attribués ayant un déplacement chimique dans la zone des protons amides. Ces informations suggèrent que Tox61pT3 est constitué de deux formes en échange conformationnel. Afin de tester cette hypothèse et pour faciliter l'attribution des pics de la forme B, une expérience de  $^1\text{H}$   $^{15}\text{N}$  ROESY hors résonance a permis d'observer des pics de corrélations caractéristiques d'un échange dans une gamme de temps comprise entre 1 et 100 ms. Cette expérience s'apparente à une expérience de  $^1\text{H}$   $^{15}\text{N}$  HSQC ; elle permet de faire apparaître des pics de corrélation lorsque deux pics correspondent à deux états conformationnels différents du même résidu, et donc à deux environnements chimiques différents (Figure 41 1). Grâce à l'utilisation de ces informations et à leur recoupement avec d'autres obtenues par l'analyse du spectre  $^1\text{H}$   $^1\text{H}$  NOESY, il a été possible d'obtenir l'attribution des pics du spectre  $^1\text{H}$   $^{15}\text{N}$  HSQC correspondant à la forme B de Tox61pT3. L'annexe 4 présente l'attribution complète de ce spectre.

#### 1.3.5.2 Valeurs des nOe. hétéronucléaire ( $^1\text{H} \rightarrow ^{15}\text{N}$ ) des formes A et B de Tox61pT3

A partir de l'attribution du spectre  $^1\text{H}$   $^{15}\text{N}$  HSQC de Tox61pT3 et des spectres saturés et non saturés, les valeurs de nOe hétéronucléaire ( $^1\text{H} \rightarrow ^{15}\text{N}$ ) des formes A et B de Tox61pT3 ont pu être calculées (Figure 41 2).

Les valeurs de nOe hétéronucléaire ( $^1\text{H} \rightarrow ^{15}\text{N}$ ) permettent de détecter l'existence de mouvements du vecteur azote amide - proton amide pour une échelle de temps de l'ordre de la picoseconde à la nanoseconde. Un résidu dont la valeur de nOe hétéronucléaire est de l'ordre de 0,8 est impliqué dans une structure. Des valeurs faibles (proches de 0 ou négatives) de nOe hétéronucléaire indiquent l'absence de structure. Les valeurs de nOe hétéronucléaire de la forme A de Tox61 sont comprises entre 0,6 et 0,8 pour les résidus des deux premiers doigts et entre 0,4 et 0,6 pour la partie C-terminale. Ces valeurs indiquent que dans l'ensemble cette forme est structurée. Dans le cas de la forme B de Tox61pT3, les résidus compris entre la cystéine 3 et l'alanine 42 (résidu muté) ont des valeurs de nOe hétéronucléaire similaires à celles de la forme A. A l'inverse, les résidus compris entre la thréonine 44 et l'extrémité C-terminale ont des valeurs nulles ou négatives.

#### 1.3.5.3 Conclusion

Tox61pT3 forme B correspond donc à une molécule structurée dans les deux premiers doigts, mais dont toute la partie C-terminale (le doigt 3 et la boucle C-terminale, qui dans la protéine native vient fermer la structure) est ici dépliée. Cette forme B est en échange permanent avec la forme A qui est quant à elle structurée comme la protéine native.

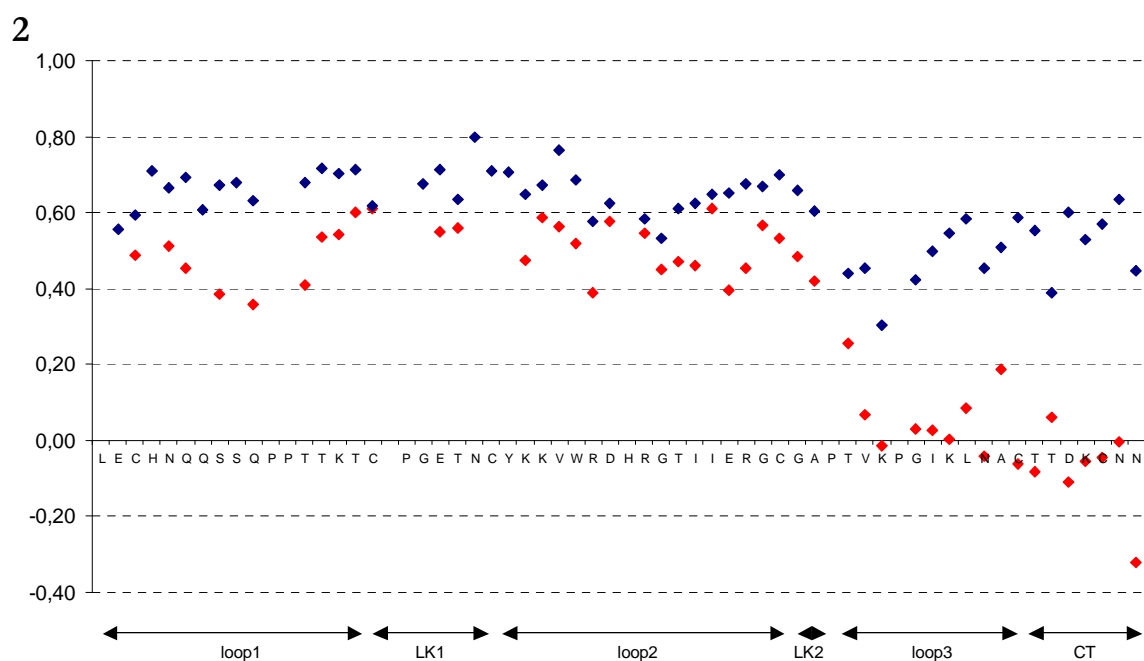
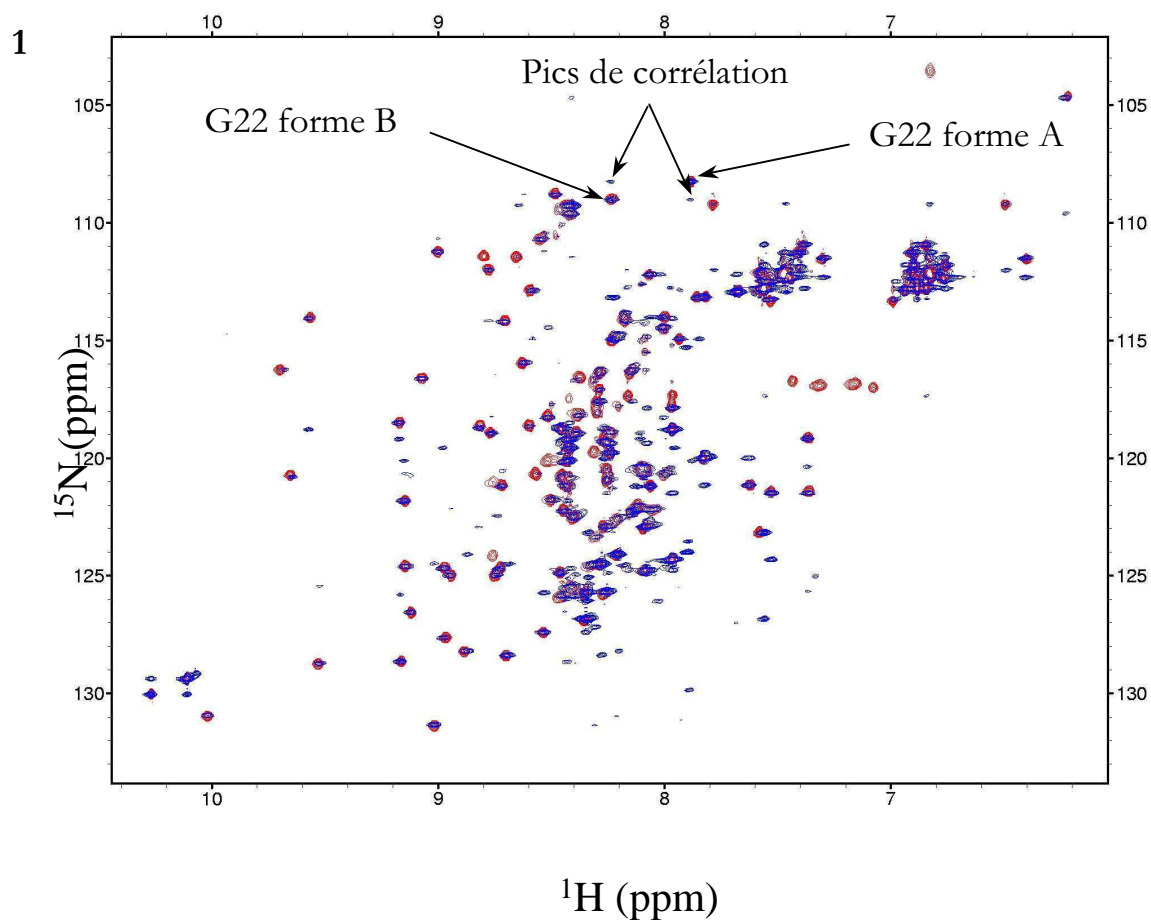


Figure 41 : 1 : Superposition du spectre  $^1\text{H}$   $^{15}\text{N}$  HSQC (en rouge) et du spectre  $^1\text{H}$   $^{15}\text{N}$  ROESY d'échange (en bleu) de Tox61pT3. 2 : nOe hétéronucléaire de Tox61pT3 forme A (en bleu), forme B (en rouge).

## 2 VECTORISATION DU REPLIEMENT D'UNE PROTEINE *IN VITRO*

Comme cela a été rappelé en introduction, le chemin de repliement cotraductionnel d'une protéine peut être très différent du chemin de repliement *in vitro* du fait de son caractère vectoriel. L'objectif de cette partie est de vectoriser le repliement d'une protéine *in vitro* (en l'occurrence la toxine  $\alpha$  de *Naja nigricollis*) en utilisant des anticorps. La stratégie consiste à utiliser des anticorps spécifiques de différentes portions de la séquence de la forme dépliée de la toxine pour la maintenir dans cet état déplié en présence d'un tampon permettant le repliement (Figure 19 présentée en partie 7.2 de l'introduction). L'inhibition du repliement serait alors levée récursivement depuis l'extrémité N-terminal vers l'extrémité C-terminale pour imposer la vectorisation. On obtiendrait ainsi une vectorisation du repliement d'une protéine monodomaine *in vitro*. Le but serait d'ensuite comparer les vitesses de repliement obtenues avec ou sans vectorisation. Une accélération est attendue lorsque le repliement est vectorisé.

Dans ce contexte, les anticorps que l'on souhaite obtenir doivent répondre au cahier des charges suivant :

- Etre spécifiques d'une partie de la séquence de la protéine dans son état réduit.
- Une fois liés, ils doivent empêcher la fermeture des ponts et donc être capables d'inhiber le repliement oxydant.
- La vitesse de dissociation de ces anticorps doit être grande par rapport à la vitesse de repliement de Tox62 pour que le démarrage du repliement dans la partie concernée soit synchronisé dans toute la population de protéines étudiées.
- Enfin, il nous faut montrer qu'il est possible de lier plusieurs de ces anticorps sur la même protéine réduite.

### 2.1 Production de solutions polyclonales d'anticorps

#### 2.1.1 Conception des peptides épitopes

L'objectif est de choisir des sections de séquence recouvrant la majeure partie de Tox62. Pour concevoir ces peptides nous avons choisi une longueur de séquence de 10 à 15 résidus qui correspond à une taille légèrement supérieure à la longueur habituellement observée pour les épitopes continus qui est comprise entre 6 et 10 résidus ((Van Regenmortel, 2001)). Les cystéines de la toxine sont remplacées dans les peptides par des sérines afin d'éviter la formation de dimères lorsqu'ils sont en solution. De plus, la substitution par une sérine doit permettre de mimer la présence d'une cystéine réduite. La seule différence entre une sérine et une cystéine libre est le remplacement de la fonction thiol par une fonction alcool. Les anticorps induits par ces peptides doivent reconnaître la forme dénaturée de la toxine ; ces peptides ont donc été choisis de sorte que leur séquence ne recouvre pas d'élément de structure secondaire qui pourraient avoir tendance à se structurer (comme deux brins d'un même feuillet  $\beta$ ). En effet, l'immunisation par un peptide étant capable d'acquérir une structure intrinsèquement conduirait à la production d'anticorps reconnaissant cette structure et non la forme dépliée. Toujours pour éviter une structuration proche de la forme oxydée de la protéine, le peptide ne doit pas recouvrir deux cystéines d'un même pont.



### 2.1.1.1 Les anticorps obtenus

Le premier jeu de peptides utilisés pour produire nos anticorps sont ceux présentés sur la Figure 42. Pour chacun des anticorps reçus, une expérience de titrage en ELISA a été réalisée afin de vérifier l'existence d'une liaison avec l'antigène polyS (Figure 43).

Les anticorps anti pTx1 ont soit une affinité très faible, soit sont en très faible quantité. Quelle que soit la raison de ce faible titre, cette solution d'anticorps ne sera pas utilisée pour la suite de la caractérisation. Les sérums obtenus à partir du peptide pTx4 n'ont montré aucune activité de liaison spécifique.

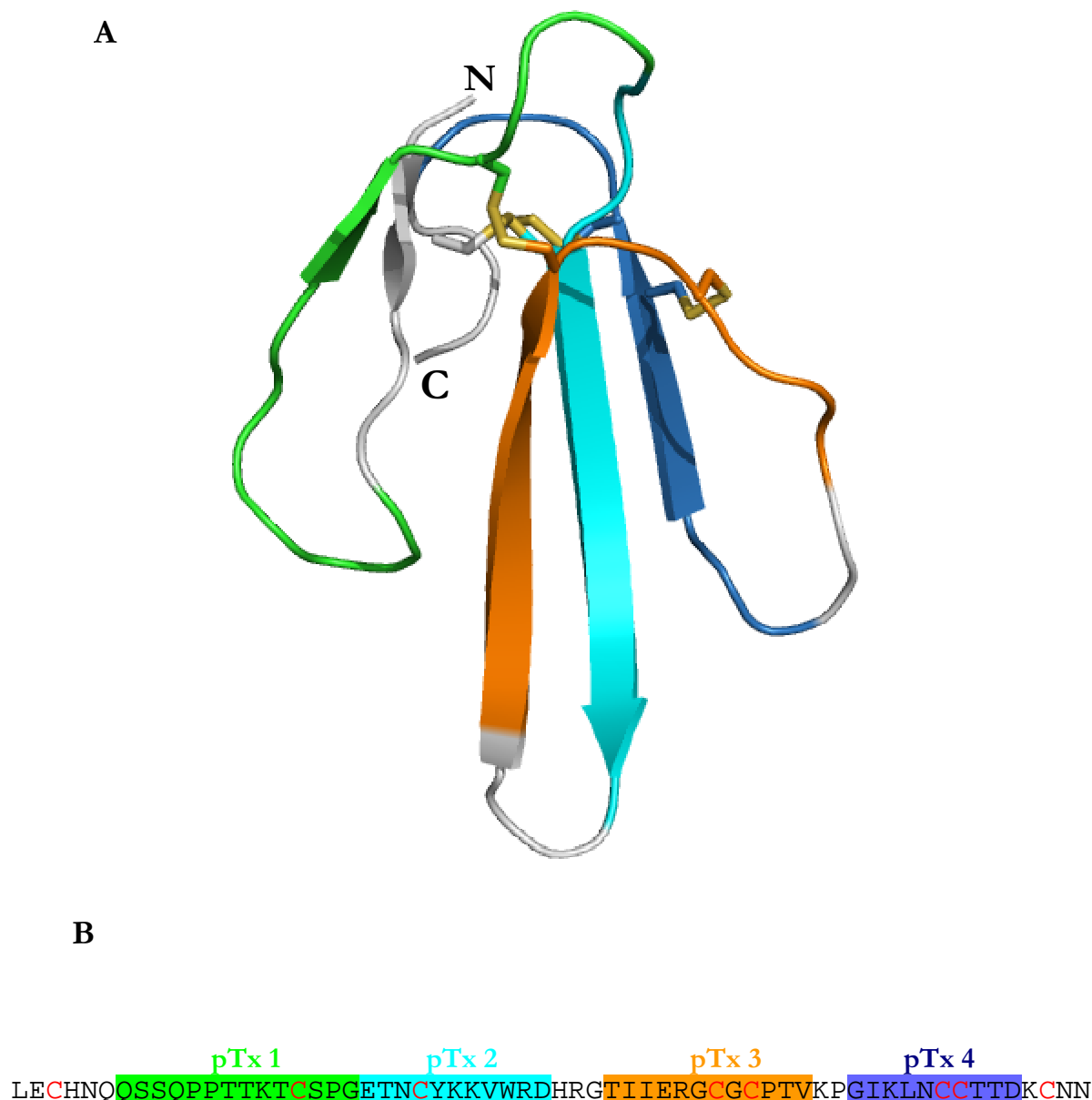
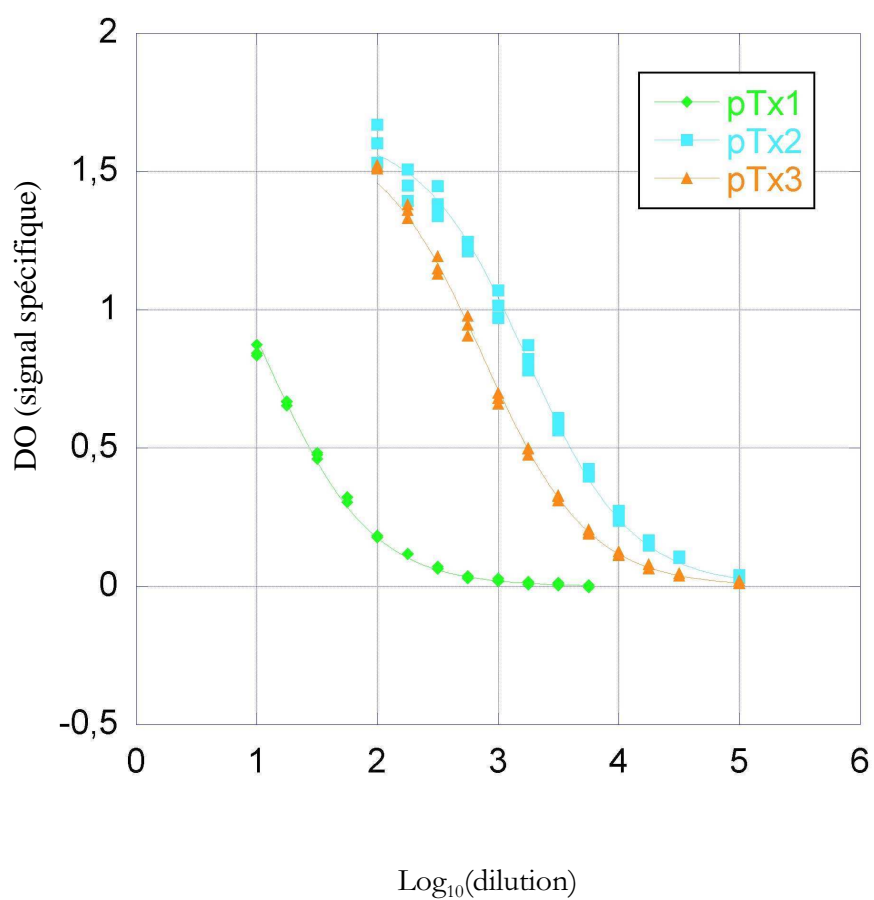


Figure 42 : A , Structure cristallographique de Tox62. Les couleurs vert, cyan, orange et bleu correspondent aux épitoopes des anticorps pTx1, pTx2, pTx3 et pTx4 respectivement. B , séquence des peptides épitoopes.

A



B

	acpTx1	acpTx2	acpTx3
Titre moyen :	<b>10</b>	<b>1900</b>	<b>500</b>

Figure 43 : A , Courbes de titrage obtenues pour les solutions brutes des anticorps anti pTx1, pTx2 et pTx3 avec polySC-biot (lié à la streptavidine immobilisée sur la plaque). B , Titres moyens obtenus avec 4 expériences de titrage par liaison sur polySC-biot immobilisé.

Ne disposant que de deux anticorps (acpTx2 et acpTx3), une autre série de peptides a été utilisée pour produire une deuxième série d'anticorps. Afin d'augmenter les chances d'obtenir un anticorps spécifique de la partie C-terminale, indispensable pour réaliser la vectorisation, la séquence de Tox62 a été rallongée par un peptide connu pour être immunogène. La séquence Etag a été choisie ((Lampugnani et al., 2006)).

**A**

pTx 1Bis pTx 3bis pTx 4Bis  
 LECHNQSSQPPTTKT CSPGETN CYKKVWRDHRGTIIERGCG PTVKPGIKLNCCTTDKCNN

**B**

pTx ETag  
 LECHNQSSQPPTTKT CSPGETN CYKKVWRDHRGTIIERGCG PTVKPGIKLNCCTTDKCNNGAPVPYPD

Figure 44 : A : Séquence des peptides utilisés pour la deuxième série d'immunisation. B : Séquence du variant Tox62Etag et du peptide pTxEtag

Le peptide épitope pTxEtag contient la séquence de Etag (GAPVPYPD) mais chevauche également l'extrémité C-terminale de Tox62 (DKCNN). Comme cela a été montré en partie 5.3.2.2 de l'introduction, l'asparagine 61 joue un rôle particulier dans la structure des protéines trois doigts. Il semble nécessaire qu'un anticorps devant bloquer le repliement de la partie C-terminale recouvre ce résidu. D'autre part cet anticorps devait recouvrir la cystéine 60 et donc empêcher la formation du pont 55-60.

Lorsque l'anticorps anti pTxEtag est utilisé, la protéine Tox62Etag (Figure 44 B) devient le modèle d'étude. Pour que nos connaissances concernant le repliement de Tox62 soient utilisables pour interpréter les résultats obtenus avec Tox62Etag, le repliement de ces deux molécules a été comparé (Figure 45).

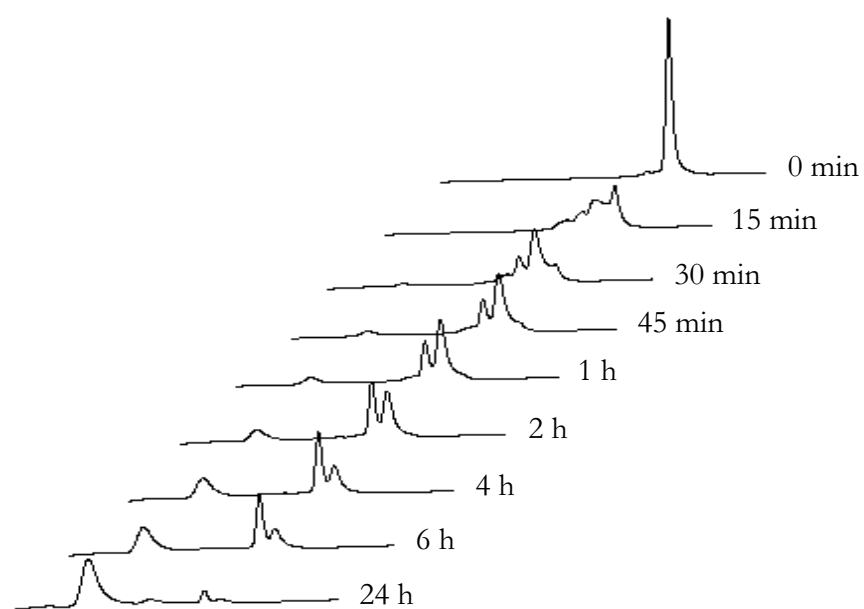
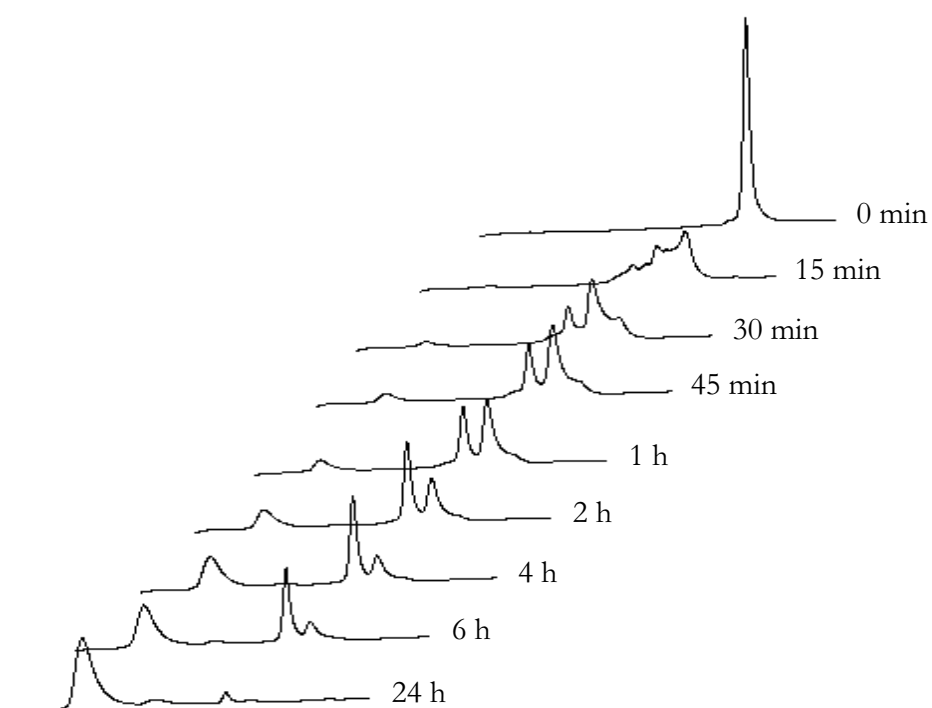


Figure 45 : Comparaison des repliements de Tox62 (1) et Tox62ETag (2) aux temps longs, suivis en HPLC-RP.

La comparaison des profils chromatographiques obtenus pour le repliement de Tox62 et Tox62Etag montre que le même nombre d'intermédiaires apparaît et que leur évolution est identique (Figure 45). Tox62Etag pourra donc être utilisé comme modèle du repliement de Tox62.

	acpTx2	acpTx3	acpTx1bis	acpTx4bis	acpTxETag
Titres :	<b>560</b>	<b>230</b>	<b>10</b>	<b>0</b>	<b>65</b>

Tableau 10 : Titres obtenus en ELISA (polySETag directement immobilisé) pour la deuxième série d'anticorps.

AcpTx2 et AcpTx3 sont ici utilisés comme référence. Leurs titres sont respectivement 3,4 et 2,2 fois inférieurs à ceux présentés sur la Figure 43 B. Cette différence peut être expliquée soit par le changement d'antigène soit par le changement de mode d'immobilisation. En effet, dans le cas du Tableau 10, l'antigène polySETag a été directement immobilisé sur la plaque alors que pour la Figure 43, l'antigène polySC-Biot est lié par sa biotine C-terminale à la streptavidine immobilisée sur la plaque. Dans le premier cas les antigènes sont adsorbés dans des configurations aléatoires sur la plaque ; dans le deuxième cas, l'antigène est immobilisé par la partie C-terminale et la partie N-terminale est orientée vers le solvant. Les autres valeurs de titre du Tableau 10 doivent donc être comparées à celles de acpTx2 et acpTx3 dans ce même tableau.

Enfin une troisième série d'anticorps a été commandée à la société PROTEOGENIX.

	acpTx2	acpTx3	acpTx3Ter	acpTxETag'
Titres :	<b>560</b>	<b>230</b>	<b>&lt;&lt; 100</b>	<b>15000</b>

Tableau 11 : Titres obtenus en ELISA (polySETag directement immobilisé) pour les anticorps fournis par la société PROTEOGENIX

### 2.1.1.2 Conclusion

Pour la suite de l'étude seuls les anticorps disposant d'un titre supérieur à 50 ont été retenus. Seuls les anticorps anti pTx2, anti pTx3, anti pTx3Ter, anti pTxETag et anti pTxETag' seront étudiés.

Il faut souligner que les anticorps anti pTx2, anti pTx3 et anti pTxETag ont été fournis par l'équipe de Marcoules. Une précipitation à l'acide caprylique a été effectuée avant l'envoi.

Anti pTx3ter et anti pTxETag' nous ont été fournis par la société PROTEOGENIX, sous forme d'un sérum. Pour obtenir une solution d'immunoglobuline G, nous avons purifié ces sérums sur colonne d'affinité à la protéine A (voir partie 4.2 des matériels et méthodes). Les caractérisations ultérieures correspondent aux solutions purifiées.

Enfin, la solution d'anticorps anti pTxETag fournie par Marcoules ayant un titre de 65 a été purifiée par chromatographie d'échange d'ions pour être ensuite concentrée sur tube Vivaspin. Ce sont les anticorps de cette solution concentrée que l'on appellera par la suite acpTxETag.

### 2.1.2 Mesure de la concentration en anticorps actifs, étude en Biacore

Dans l'optique de réaliser la vectorisation du repliement, il est nécessaire de se placer en excès en anticorps actifs par rapport à la toxine réduite pour que toutes les molécules de toxine subissent l'action des anticorps. Nous avons donc besoin de déterminer la concentration en anticorps actifs dans les solutions d'anticorps dont nous disposons. Ces solutions d'anticorps sont toutes des mélanges polyclonaux. Il convient de noter qu'une solution polyclonale d'anticorps contient des anticorps pouvant présenter des affinités hétérogènes, en conséquence il est probable que toutes les molécules de toxine ne se lient pas de la même façon aux anticorps. On se propose ici de déterminer une concentration équivalente en assimilant le mélange polyclonal à une solution pure d'anticorps d'affinité homogène (Tableau 12). Par la suite, un excès d'anticorps par rapport à la toxine sera utilisé afin de palier à l'imprécision de cette mesure.

	acpTx2	acpTx3	acpTx3ter	acpTxETag	acpTxETag'
Concentration en anticorps spécifiques en nM	<b>280</b>	<b>110</b>	<b>55 000</b>	<b>760</b>	<b>2 700</b>

Tableau 12 : Tableau récapitulatif des valeurs de concentration en anticorps spécifiques

## 2.2 Caractérisation des anticorps obtenus

Cette partie a pour objectif de vérifier que les solutions d'anticorps obtenues répondent au cahier des charges fixé au début de cette partie concernant la vectorisation du repliement.

### 2.2.1 Spécificité

Cette étude a été menée uniquement sur les anticorps anti pTx2, pTx3 et pTxETag. Dans le contexte d'une vectorisation du repliement par des anticorps, deux questions se posent concernant la spécificité de chacun de ces anticorps :

1. L'anticorps reconnaît-il l'épitope au sein de la protéine polyS ?
2. L'anticorps reconnaît-il préférentiellement la forme réduite de Tox62 par rapport à la forme oxydée?

Ces questions ont été étudiées par des expériences d'ELISA (Enzyme-linked immunosorbent assay). Cette méthode a été choisie car elle est très sensible et nécessite donc une faible quantité d'anticorps (notre source d'anticorps était limitée). De plus, elle permet de contrôler aisément les conditions dans lesquelles a lieu la liaison.

Des expériences de compétition à l'équilibre ont été menées. L'objectif de ce type d'expérience est de comparer l'affinité d'un anticorps pour deux antigènes. Un des antigènes sert de référence et est immobilisé dans un puits de la plaque d'ELISA. Le deuxième antigène et l'anticorps sont incubés dans ce même puits. L'anticorps entre alors en compétition pour la liaison à l'antigène fixé et l'antigène libre. Après lavage et révélation, un signal proportionnel à la quantité d'anticorps encore lié à l'antigène immobilisé est obtenu. Les concentrations en antigène immobilisé et en anticorps sont maintenues constantes, alors que la concentration en antigène libre est croissante. Le but est de déterminer la concentration en antigène libre nécessaire pour déplacer 50 % des anticorps ( $IC_{50}$ ). Il est ensuite possible de comparer les valeurs d' $IC_{50}$  obtenues pour déterminer l'affinité relative de l'anticorps pour les différents antigènes compétiteurs.

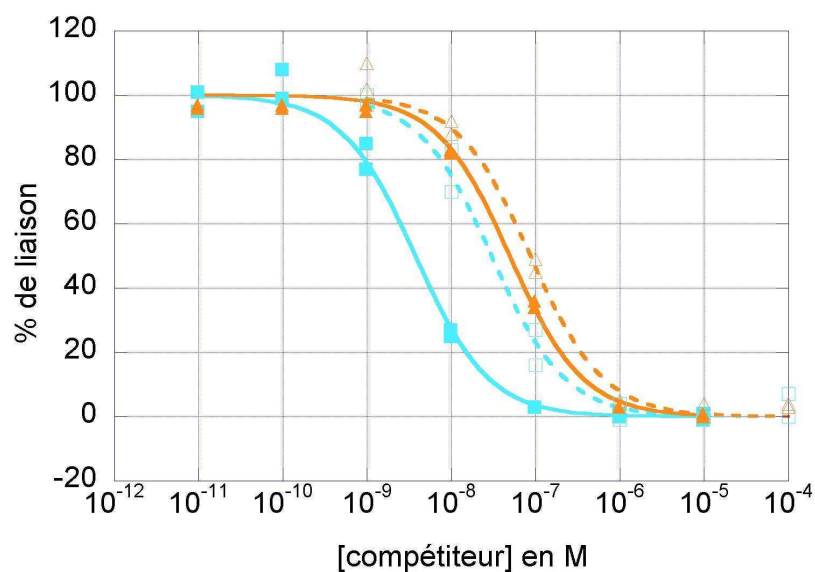
#### 2.2.1.1 Question 1 : comparaison polyS/peptide

Les courbes de titrage en ELISA sont présentées en Figure 46 1 et 2. Les  $IC_{50}$  correspondant figurent dans le tableau associé (Figure 46 3).

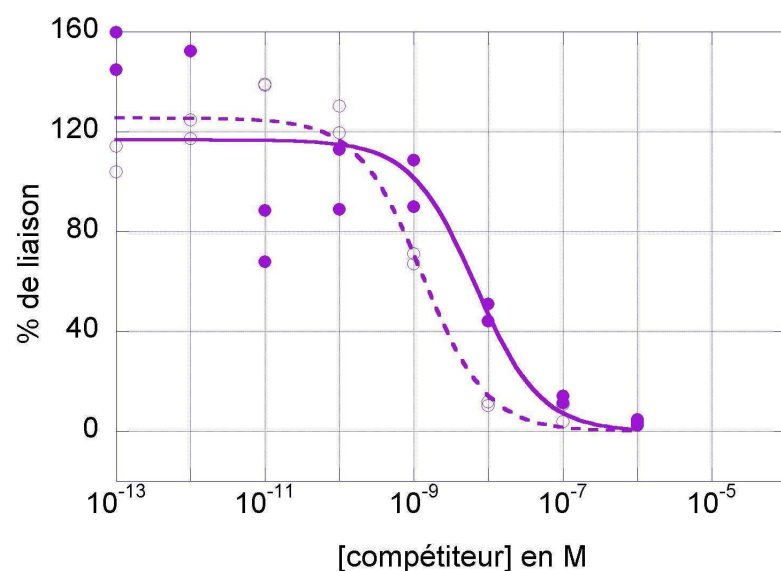
L'anticorps anti pTx3 a une affinité identique pour polyS et le peptide épitope. Cet anticorps reconnaît donc indifféremment son épitope libre ou inclus au sein de polyS.

A l'inverse pour acpTx2 une différence d' $IC_{50}$  d'un facteur 10 est observé. L'anticorps anti pTx2 reconnaît préférentiellement l'épitope libre par rapport à l'épitope inclus dans le contexte polyS. Cette différence peut être liée à un problème de présentation de l'antigène. Cependant l'affinité de l'anticorps anti pTx2 n'est pas dramatiquement affectée puisque son  $IC_{50}$  est de 50nM. Son affinité est donc considérée comme suffisante pour continuer le processus de caractérisation. AcpTxETag fait preuve d'une affinité légèrement meilleure pour polySETag que pour son peptide. Le facteur 7 observé peut être attribué à l'imprécision de la méthode. Toutefois son affinité pour polySETag est suffisante pour pouvoir poursuivre le processus de caractérisation. On peut noter que tous les anticorps se situent dans une même gamme d'affinité.

1



2



3

	acpTx2	acpTx3	acpTxETag
IC <sub>50</sub> vs peptide (nM)	4	30	7
IC <sub>50</sub> vs polyS (nM)	50	87	1

Figure 46 : Courbes de compétition à l'équilibre pour les anticorps anti pTx2 (cyan) et anti pTx3 (orange) dirigés contre polySC-Biot immobilisé sur la plaque (1) et pour l'anticorps anti pTxETag (violet) dirigé contre polySETagC-Biot (2). La courbe pleine correspond à la compétition avec le peptide, la courbe en pointillés à celle avec polyS (ou polySETag à droite). 3 : Tableau des IC<sub>50</sub> obtenus.



### 2.2.1.2 Question 2 : comparaison toxine réduite / toxine oxydée

Dans les expériences d'ELISA de compétition contre la forme réduite, le milieu est supplémenté en DTT réduit pour maintenir la toxine dans son état réduit.

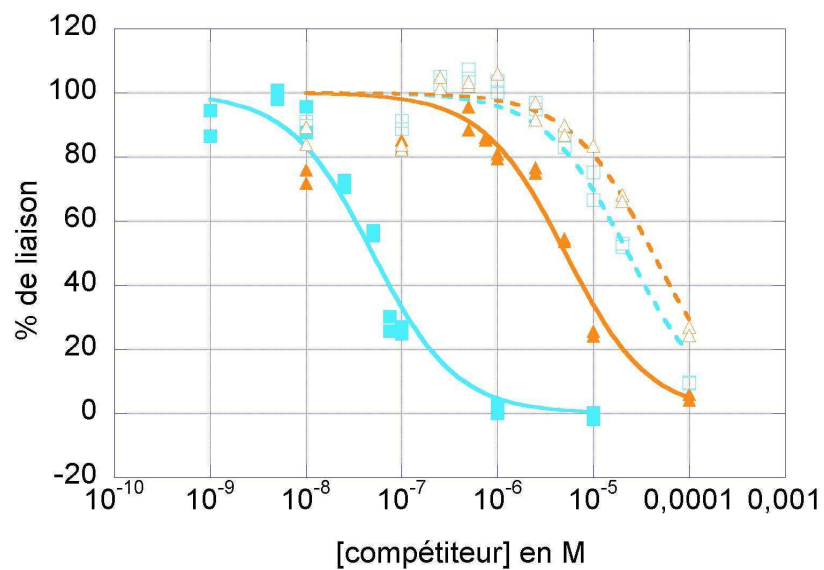
En comparant les résultats des compétitions contre les formes réduites avec ceux obtenus par compétition avec polyS et polySEtag (Figure 47), il est possible d'évaluer la perte d'affinité inhérente au remplacement d'une sérine par une cystéine libre. Pour acpTx2 et acpTxETag, aucune perte d'affinité n'est observée puisque les titres sont identiques (Figure 47 1 et 2). Pour acpTx3, l'affinité est diminuée d'un facteur 160.

D'autre part, acpTx2 reconnaît préférentiellement la forme réduite de Tox62 plutôt que sa forme oxydée. En effet les  $IC_{50}$  obtenus diffèrent d'un facteur 500 environ. L' $IC_{50}$  de la forme réduite dans le cas de acpTx3 est seulement 10 fois plus faible que celui de la forme oxydée. La faible différence d'affinité entre la forme réduite et la forme oxydée est probablement imputable dans ce cas à une faible affinité pour la forme réduite dans l'absolu.

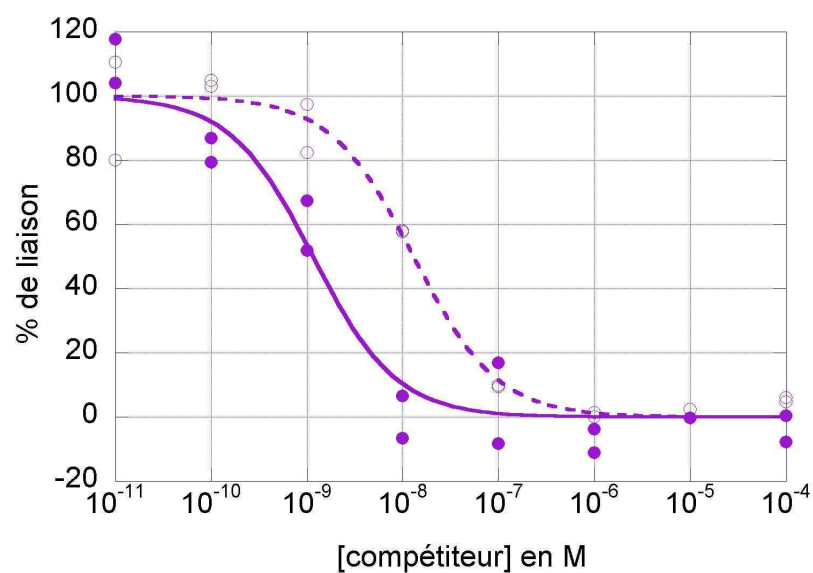
Enfin concernant acpTxETag, comme dans le cas de acpTx3, le rapport des  $IC_{50}$  entre forme réduite et forme oxydée n'est que de 10. Ce faible écart est dû cette fois-ci à une forte affinité de l'anticorps pour la forme oxydée de Tox62ETag (soit 10nM contre 1nM pour polySEtag). Ce résultat peut s'expliquer par le fait que le peptide ETag, ajouté à l'extrémité C-terminale de Tox62 est peut-être accessible que la protéine soit réduite et dénaturée ou oxydée et structurée.

Ces résultats démontrent que tous les anticorps étudiés répondent à la nécessité de reconnaître préférentiellement la forme réduite par rapport à la forme oxydée et donc la validité du principe utilisé pour les produire.

1



2



3

	acpTx2	acpTx3	acpTxETag
IC <sub>50</sub> vs Toxine réduite (nM)	<b>50</b>	<b>5 000</b>	<b>1</b>
IC <sub>50</sub> vs Toxine oxydée (nM)	<b>23 000</b>	<b>42 000</b>	<b>13</b>

Figure 47 : Courbes de compétition à l'équilibre pour les anticorps anti pTx2 (cyan) et anti pTx3 (orange) dirigés contre polySC-Biot immobilisé sur la plaque (1) et pour l'anticorps anti pTxETag (violet) dirigé contre polySETagC-Biot(N-Ter) (2). La courbe pleine correspond à la compétition avec la toxine Tox62 réduite, la courbe en pointillés à celle avec la toxine Tox62 oxydée (à droite la toxine utilisée est Tox62ETag). 3 : Tableau des IC<sub>50</sub> obtenus.

### 2.2.2 Vitesse de dissociation moyenne

Dans l'optique de réaliser la vectorisation du repliement de la toxine, il faut que la vitesse de dissociation des anticorps soit grande par rapport au repliement de la toxine. En effet si l'objectif est d'observer une accélération du repliement lors de la vectorisation, il est souhaitable que le temps de dissociation ne masque pas l'éventuel gain d'efficacité. De plus, quel que soit l'objectif, un autre problème se pose. Lors de la vectorisation, il est nécessaire de dissocier les anticorps à un instant donné. Si la dissociation est lente comparée à la vitesse de repliement, la population de toxines risque de ne pas être homogène, certaines protéines se repliant alors que d'autres seraient encore liées aux anticorps. Il serait alors difficile d'analyser les résultats.

Les cinétiques de dissociation de acpTx2 et acpTx3 ont été suivies par Radio Immuno Assay (voir partie 4.3.4 des résultats et méthodes). La courbe suivante présente la dissociation de acpTx2 de Tox62 réduite. La courbe de modélisation tracée correspond à un mécanisme de dissociation du premier ordre.

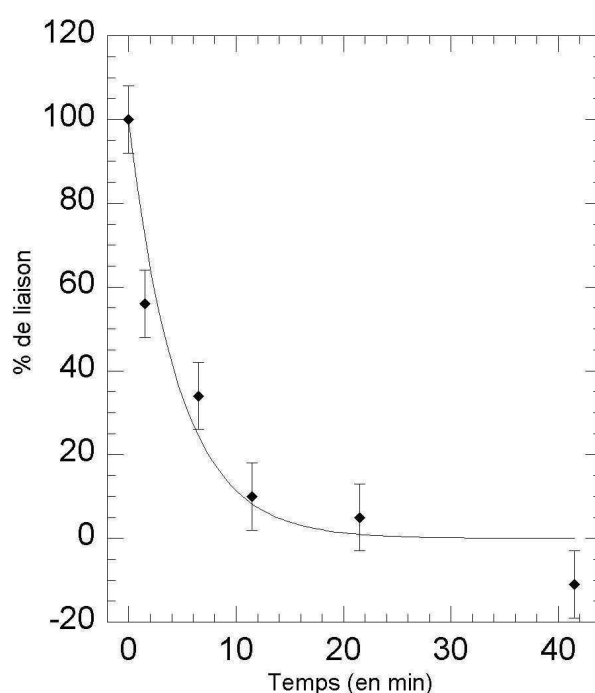


Figure 48 : Cinétique de dissociation du complexe acpTx2 / Tox62 red. (coefficient de corrélation : 0,97)

La modélisation des résultats obtenus lors de la cinétique de dissociation du complexe acpTx2 / Tox62 red (Figure 48) donne un temps de demi-dissociation de 3,5 minutes. Ceci implique qu'en 10 minutes 90% des toxines réduites sont dissociées. Ce temps est 23 fois inférieur au temps de demi-repliement de Tox62 qui est d'un peu moins de 4 heures. Il faut remarquer que cette valeur correspond à une vitesse de dissociation moyenne puisque acpTx2 est un mélange polyclonal d'anticorps.

Pour acpTx3, aucun signal spécifique n'a pu être obtenu. L'expérience a été reproduite sans succès. Deux interprétations sont possibles, soit le  $K_d$  (constante d'équilibre de dissociation) de acpTx3 est trop important pour qu'une proportion détectable de toxine soit liée dans les conditions de l'expérience à savoir 100nM de Tox62  $^3H$ , 10nM d'anticorps spécifiques ; soit la cinétique de dissociation est tellement rapide que pendant la phase de traitement des prélèvements les complexes se dissocient, auquel cas le temps de demi-dissociation serait inférieur à la minute.

### 2.2.3 Liaison de plusieurs anticorps sur un seul antigène

#### 2.2.3.1 Gel filtration et immunoblot

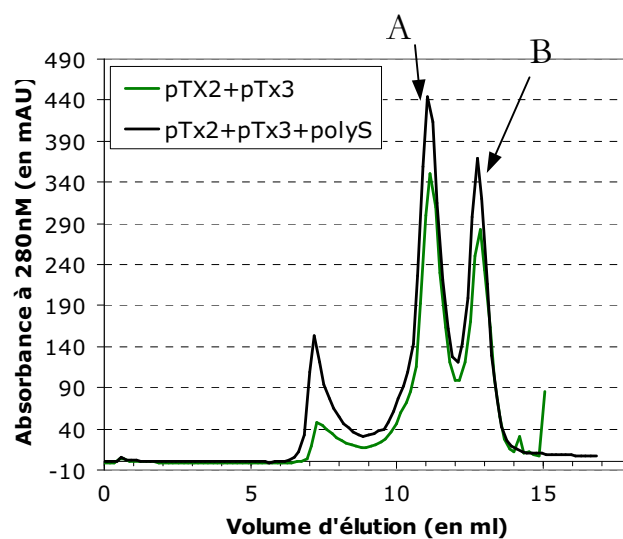
- Gel filtration

Le but de cette étude est d'observer le complexe tripartite entre acpTx2/PolyS/acpTx3. Pour obtenir une quantité maximale de complexes ternaires, des conditions d'excès en anticorps par rapport à polyS sont utilisées. La concentration en anticorps anti pTx2 et anti pTx3 a été fixée à 70 nM (soit 10 pmol d'anticorps actif), contre 20 nM pour polyS (soit 3 pmol). Le mélange est incubé en tampon PBS pendant une heure à 37°C.

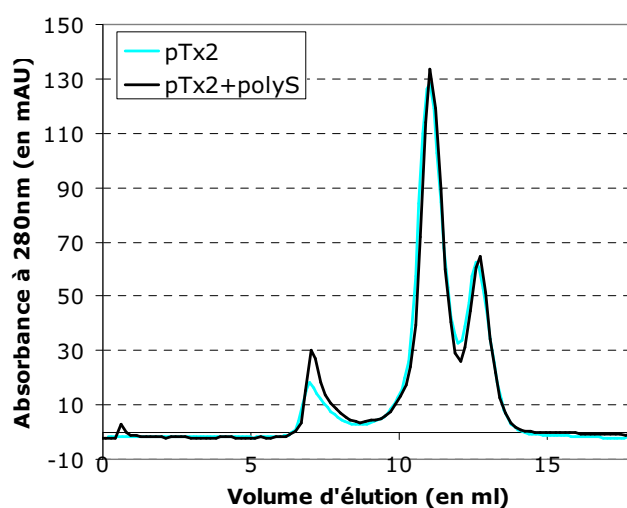
Le chromatogramme de gel filtration du mélange contenant les trois partenaires est présenté en Figure 49 1. Comme expérience contrôle, des mélanges contenant uniquement acpTx2 et pTx2 (Figure 49 2) et uniquement acpTx3 et pTx3 (Figure 49 3) ont été analysés dans les mêmes conditions en gel filtration.

Le chromatogramme d'analyse de gel filtration du mélange tripartite (Figure 49 1) contient un pic A ayant un volume d'élution de 11,1 mL correspondant à une masse moléculaire de 160 000 Da. Cette masse moléculaire correspond à celle des anticorps. Le volume d'élution du pic B est de 12,7 mL ce qui correspond à une masse moléculaire de 61000 Da, ce pic est probablement composé d'albumine résiduelle. La comparaison de l'analyse du mélange contenant les anticorps anti pTx2 et anti pTx3 d'une part et du même mélange dans lequel la protéine polyS a été ajoutée d'autre part ne permet pas d'observer l'apparition d'un pic supplémentaire correspondant au complexe ternaire. Ce complexe devrait avoir une masse moléculaire d'environ 300 000 Da et donc sortir après l'élution de 10 mL de tampon. Un léger épaulement existe pour ce volume d'élution mais il est également présent dans le chromatogramme correspondant à l'anticorps anti pTx3 seul (Figure 49 3). Il faut noter que les proportions d'anticorps spécifiques dans ces mélanges polyclonaux sont de 0,6 et 0,2% pour acpTx2 et acpTx3 respectivement. Il semble difficile de mettre en évidence le complexe ternaire par cette méthode puisque le pic à identifier serait masqué par celui des anticorps libres.

1



2



3

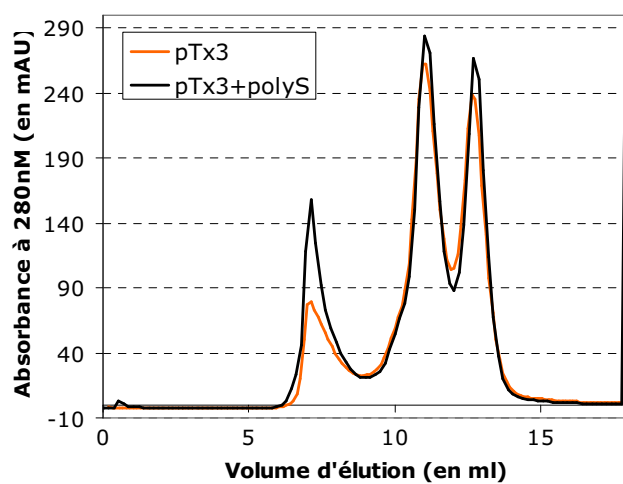


Figure 49 : Chromatogrammes de gel filtration

Pour obtenir une preuve de l'existence du complexe ternaire, les fractions de gel filtration correspondant à la masse moléculaire attendue de ce complexe (6 fractions de 200µL entre 10 et 11,2mL équivalent à des masses moléculaires de 300 kDa à 85 kDa) ont été analysées en immunoblot (Figure 50).

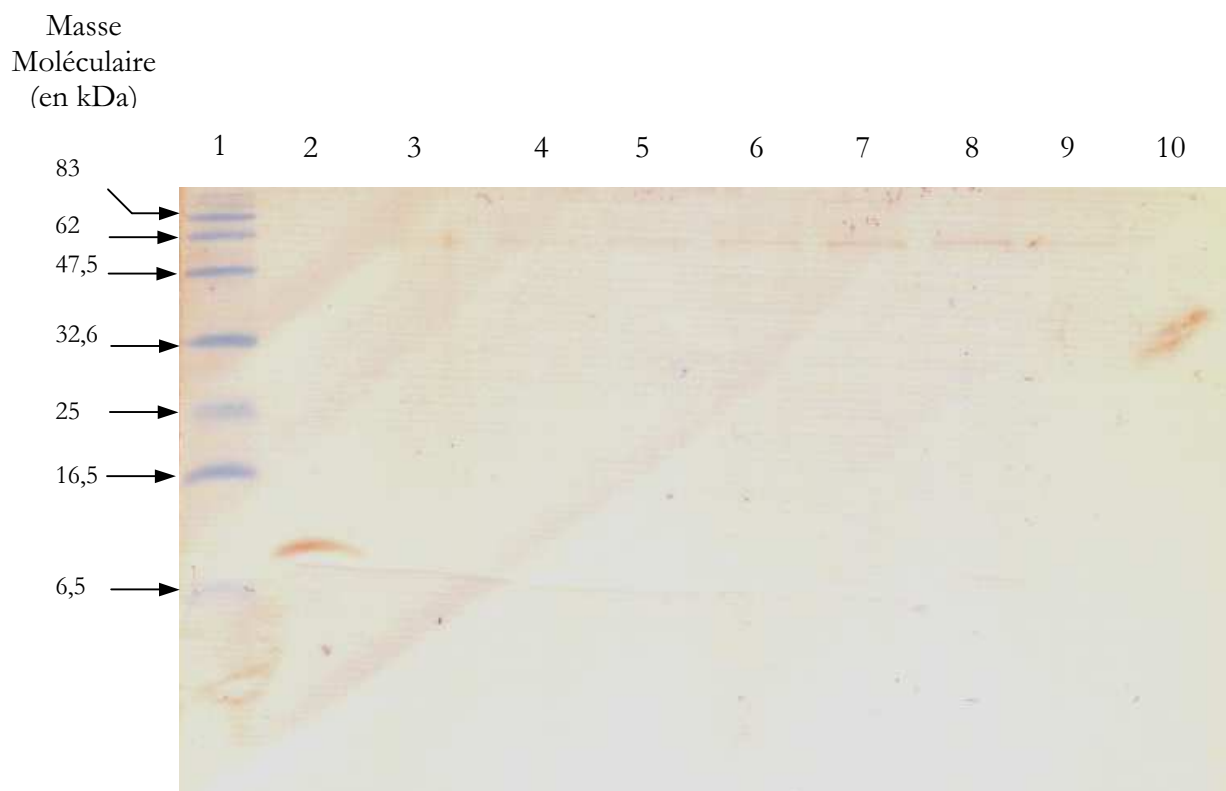


Figure 50 : Immunoblot des fractions de gel filtration. 1 : Marqueur de poids moléculaire, 2 : 4ng de polyS, 3 à 9 : fractions de gel filtration du mélange acpTx2 + acpTx3 + polyS de 10 à 12,4mL tous les 200µL

Une bande correspondant à la masse moléculaire de la toxine polyS (soit 6,8 kDa) apparaît en ligne 2. Aucune des lignes de migration des fractions de gel filtration analysées ne contient cette bande. Des expériences de mise au point ont démontré que par cette méthode il est possible de détecter une quantité de 0,5 ng de polyS. Le mélange de formation du complexe contient 10 pmol de acpTx2, 10 pmol de acpTx3 et 3 pmol de polyS. Si 100% des molécules de polyS sont impliquées dans la liaison tripartite, on doit récupérer 3 pmol de polyS (soit 20ng) dans les fractions de gel filtration correspondant à un poids moléculaire de 300 kDa. On fait l'hypothèse que le complexe s'étale sur 2 mL de temps de rétention. Chaque fraction de 200 µL contient alors théoriquement 2 ng de toxine polyS. Ceci est quatre fois supérieur au seuil de détection de la méthode.

Cette méthode ne nous a pas permis de mettre en évidence le complexe tripartite.

### 2.2.3.2 Etude en Biacore de la liaison de deux anticorps sur polyS immobilisé

Cette étude a été réalisée en Biacore (voir partie 4.3.2 des matériels et méthodes qui détaille la méthodologie générale liée au Biacore). L'objectif est de lier un anticorps sur polyS immobilisé afin d'atteindre un signal proche de la saturation, puis d'injecter dans un deuxième temps le deuxième anticorps et d'observer si la liaison obtenue est indépendante de la présence du premier anticorps. Pour atteindre le plus rapidement possible la saturation tout en consommant le moins d'anticorps possible, une puce sur laquelle a été immobilisée une faible quantité de polySC-biot (40RU) et un faible débit d'injection (5µL/min) ont été choisis.

Deux expériences ont été effectuées (Figure 51). La première (en rouge) est composée de 4 injections successives (sans phase de régénération) de 650µL. Dans un premier temps on injecte l'anticorps anti pTx2 au 40<sup>ème</sup> (1.1), puis l'anticorps anti pTx3 au 40<sup>ème</sup> (1.2), le peptide pTx3 à 0,5M (1.3) et enfin la le peptide pTx2 à 0,5M (1.4). La deuxième expérience (en vert) est également composée de 4 injections de 650µL mais cette fois-ci l'ordre est inversé : la première contient l'anticorps anti pTx3 au 40<sup>ème</sup>, la deuxième l'anticorps anti pTx2 au 40<sup>ème</sup>, la troisième le peptide pTx2 à 0,5M et la quatrième le peptide pTx3 à 0,5M.

Analysons tout d'abord les portions de sensorgramme 1.1 et 2.1 qui correspondent aux liaisons des anticorps acpTx2 et acpTx3 respectivement sur une puce où tous les antigènes sont libres. L'amplitude du signal est de 350 RU pour acpTx2 et 150 RU pour acpTx3. Cette différence s'explique par la différence de concentration en anticorps actifs utilisée d'un cas à l'autre (7nM pour acpTx2 et 3 nM). Ces conditions ont été choisies pour s'approcher le plus possible de la saturation des antigènes tout en diluant suffisamment les solutions d'anticorps dans le tampon Biacore, afin de diminuer le bruit de fond généré par la présence de fortes concentrations en anticorps non-spécifiques. Dans les deux cas, les conditions choisies sont telles que l'augmentation du signal en fonction du temps n'est pas linéaire, elle dépend de la quantité d'anticorps déjà fixée et donc en réalité de la quantité d'antigènes encore disponibles sur la puce.

Si la liaison des deux anticorps sur une molécule de polySC-Biot immobilisée était exclusive, l'amplitude de la liaison de acpTx2 (2.2) après la liaison de acpTx3, devrait être égale à la différence entre l'amplitude de acpTx2 seul (1.1) et l'amplitude de acpTx3 seul soit 200 RU. Si la liaison des deux anticorps était parfaitement indépendante, cette même amplitude devrait être égale à l'amplitude de liaison de acpTx2 seul soit 350 RU. On observe une amplitude de liaison de 275 RU et donc une situation intermédiaire. Il semble donc qu'une partie des antigènes ayant reçu l'anticorps acpTx3 soient encore disponible pour la liaison avec l'anticorps acpTx2. Lorsque l'expérience est menée dans l'ordre inverse (run 2), la même situation est observée puisque l'amplitude de signal de liaison de acpTx3 après la liaison de acpTx2 (1.2) est de 100 RU à comparer aux 150RU obtenus lorsque la puce est vierge.

Une autre interprétation peut être proposée pour expliquer ces résultats. En effet si les antigènes immobilisés étaient composés de deux populations, dont une disposerait uniquement de l'épitope de acpTx2 et l'autre uniquement de l'épitope de acpTx3, des résultats comparables à ceux présentés sur la Figure 51 seraient obtenus alors que les anticorps seraient parfaitement compétitifs. Voyons si cette hypothèse est plausible. On peut imaginer deux scénarios aboutissant à l'existence de deux populations d'antigène : une hétérogénéité dans la molécule de départ (polySC) ou une hétérogénéité dans l'immobilisation de ces molécules. Concernant la molécule de départ, il faut souligner que la purification de la molécule de polySC utilisée pour produire polySC-biot s'est terminée par une étape d'HPLC en phase inverse où un seul pic a été observé. Concernant l'immobilisation sur la puce, la stratégie utilisée impliquant la liaison spécifique streptavidine-biotine assure une immobilisation homogène par l'extrémité C-terminale de la molécule.

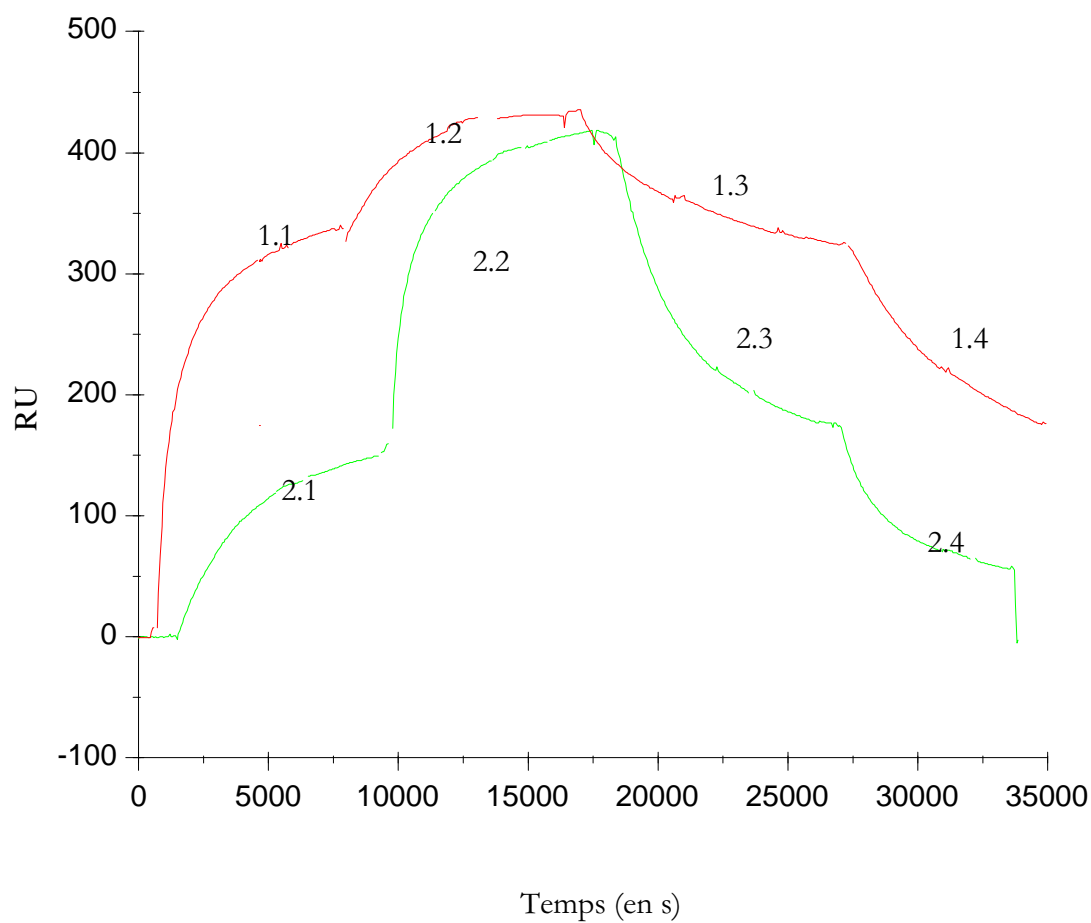


Figure 51 : Superposition des sensorgrammes obtenus pour le run 1 (en rouge) et le run 2 (en vert). Les explications sont dans le texte. Amplitude des signaux observés : 1.1 : 350 RU 1.2 : 100 RU; 2.1 : 150 RU 2.2 : 275 RU



## 2.3 Effet des anticorps sur le repliement de la toxine $\alpha$

### 2.3.1 Méthodes d'étude de la cinétique de repliement à faible concentration

Les résultats de la partie précédente ont montré que les solutions d'anticorps polyclonaux dont nous disposons ont des concentrations de l'ordre de la centaine de nanomolaires (voir Tableau 12 en partie 2.1.2 de ces résultats). La mesure de cette concentration étant approximative, il est nécessaire de se placer dans un excès d'anticorps, de cinq à dix fois. La méthode utilisée pour suivre le repliement doit permettre de déterminer l'avancement du repliement de la toxine  $\alpha$  à des concentrations de l'ordre de la dizaine de nanomolaires. Deux méthodes ont été développées.

#### 2.3.1.1 Repliement suivi par mesure de l'activité biologique

##### ■ Test d'activité biologique

Afin de déterminer l'activité biologique d'une toxine, des tests de liaison ont été développés au laboratoire par l'équipe de Denis Servent (SIMPRO CEA-Saclay). Le principe de cette expérience consiste, en une compétition entre la toxine que l'on étudie et une toxine de référence marquée radioactivement, ici la bungarotoxine  $^{125}\text{I}$ . En réalisant une gamme de concentration de Tox62 on obtient la courbe de déplacement suivante :

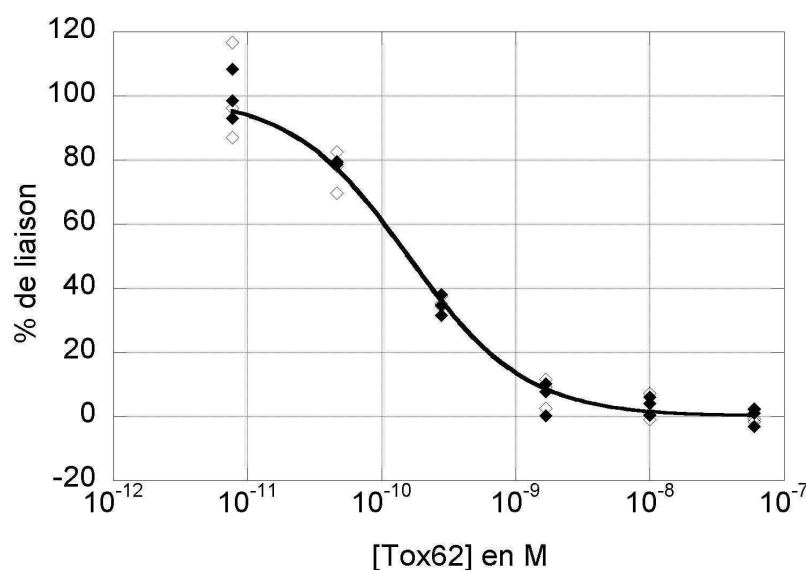


Figure 52 : expérience de compétition entre Tox62 et la bungarotoxine  $^{125}\text{I}$ . Les losanges pleins correspondent à une incubation réalisée dans le tampon PBS+BSA 0,1% alors que les losanges vides correspondent à une incubation dans le tampon de repliement.

On observe qu'entre 0,02 nM et 1 nM le signal obtenu est dépendant de la concentration en toxine (Figure 52). Cette méthode permet donc de détecter des formes de toxine biologiquement actives dans une gamme de concentration compatible avec notre objectif. Le but de l'expérience étant d'étudier la concentration en toxine active dans un mélange réactionnel de repliement il est nécessaire de vérifier que la liaison de Tox62 et de la bungarotoxine iodée ne sont pas affectées par la présence du tampon oxydoréducteur. La même gamme est obtenue dans les conditions de l'expérience de repliement. L' $\text{IC}_{50}$  de Tox62 n'est pas modifié par ces conditions (Figure 52, losange vides). Dans les deux cas il est possible de modéliser ces expériences par l'équation de l'isotherme de Langmuir.

$$S = S_0 + (S_T - S_0) / (1 + [Tox62] / IC_{50})$$

A partir de cette équation, la concentration en Tox62 est exprimée en fonction du signal observé (S), de l'IC<sub>50</sub>, du signal maximum (S<sub>T</sub>) et du signal non spécifique (S<sub>0</sub>) :

$$[Tox62] = IC_{50} * ((S_T - S_0) / (S - S_0) - 1)$$

Plusieurs expériences de compétition ont été effectuées dans ces mêmes conditions, l'IC<sub>50</sub> moyen observé était de 1,6 .10<sup>-10</sup> M.

Lorsque le signal maximum et le signal non spécifique sont connus, il est possible de calculer la concentration en forme active de Tox62 en fonction du signal observé, grâce à la formule suivante :

$$[Tox62] = 1,6 .10^{-10} * ((S_T - S_0) / (S - S_0) - 1)$$

#### ▪ Expérience de repliement

Une expérience de repliement oxydant de Tox62 est effectuée à une concentration de 12,5nM et est suivie en activité biologique. Dans ce cas le blocage est effectué par ajout d'une solution de iodoacétamide. L'objectif est de comparer la cinétique de repliement à faible concentration suivie en activité biologique à celle obtenue en HPLC avec une concentration de 15μM. Les résultats de ces deux cinétiques ne sauraient être comparés directement car la mesure d'activité donne une valeur globale de l'activité du mélange de repliement. Or, les formes C et D ayant une structure proche du natif, elles sont susceptibles d'avoir une activité biologique. Les IC<sub>50</sub> des intermédiaires C et D (bloqués à l'iodoacétamide et purifiés en chromatographie phase inverse) ont été mesurés (Tableau 13).

	Tox62	Tox62 int C	Tox62 int D
IC <sub>50</sub> en nM	<b>0,16</b>	<b>0,87</b>	<b>3,5</b>

Tableau 13 : Mesure de l'IC<sub>50</sub> des intermédiaires C et D de Tox62 bloqués à l'iodoacétamide.

Lorsque ces valeurs sont comparées à celles de Tox62 (1,6.10<sup>-10</sup> M), C et D sont respectivement 5,4 et 22 fois moins actifs que Tox62. Sachant que ces intermédiaires apparaissent très tôt et en quantité importante lors du repliement, il est indispensable de tenir compte de leur présence lors de la comparaison entre les résultats obtenus en HPLC et ceux obtenus par mesure de l'activité biologique.

Dans les conditions utilisées lors de l'expérience de repliement, c'est à dire pour une concentration en protéine comprise entre 0,1 et 1nM , le pourcentage de liaison de bungarotoxine <sup>125</sup>I est proportionnel à la concentration en intermédiaire C et en intermédiaire D. Il faut 5,4 fois plus d'intermédiaire C que de forme native pour réaliser une même inhibition et de la même façon 22 fois plus d'intermédiaire D. A partir de ce raisonnement on peut calculer une proportion équivalente de protéine active à partir des proportions des formes C, D et N obtenues en HPLC à partir de la formule suivante :

$$P_{equiv} = \%N + \%C/5,4 + \%D/22$$

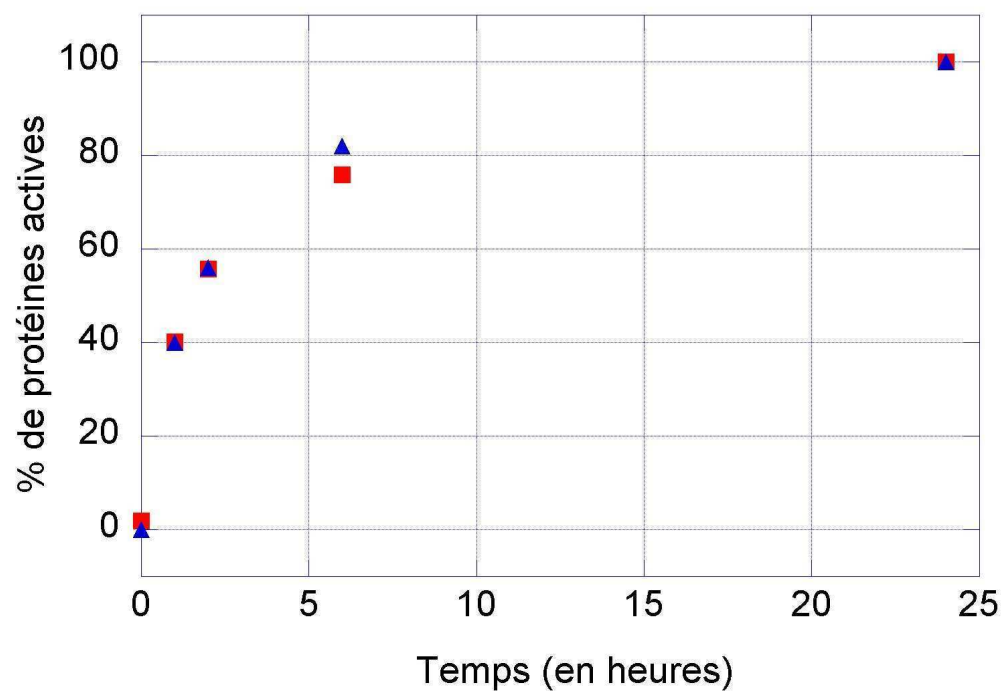


Figure 53: Cinétique de repliement de Tox62 suivie en activité biologique à 12,5nM (carrés rouges) et simulation en HPLC (triangles bleus)

#### ▪ Conclusion

Ces résultats signifient que la vitesse de repliement est identique que la concentration en toxine soit de 12,5nM ou de 15 $\mu$ M (Figure 53). Par ailleurs le suivi de l'activité biologique semble être une méthode adéquate pour étudier le repliement des toxines à une concentration de l'ordre du nanomolaire.

### 2.3.1.2 Repliement suivi par isoélectrophorèse

La méthode d'étude du repliement par mesure de l'activité biologique est certes très sensible mais ne fournit aucune information sur les formes intermédiaires qui apparaissent lors du repliement oxydant. Afin d'obtenir ce type d'information, une méthode de séparation des intermédiaires formés lors du repliement par isoélectrophorèse a été développée. Le repliement des intermédiaires dont certaines des cystéines sont libres est stoppé grâce à un agent introduisant une charge par cystéine libre (MTSBS). Cette méthode permet de séparer les espèces en fonction de leur nombre de cystéines libres.

La sensibilité nécessaire à la détection de faibles quantités de toxine a été obtenue par l'utilisation d'une toxine marquée au soufre  $^{35}\text{S}$ . Pour n'avoir à produire qu'une seule protéine marquée, nous avons utilisé une toxine capable de se lier aussi bien à acpTx2, acpTx3 et acpTxETag, c'est à dire Tox62ETag.

Les cinétiques de repliement de Tox62 et Tox62ETag sont similaires (Figure 45 de la partie 2.1.1 des résultats). Afin de connaître le comportement en isoélectrophorèse des formes apparaissant lors du repliement de Tox62ETag, les intermédiaires C et D de Tox62ETag froide ont été purifiés par HPLC phase inverse puis bloqués au MTSBS. Ces intermédiaires ont ensuite été analysés en isoélectrophorèse (Figure 54).

Les formes réduite, C, D et native de Tox62 ETag ont un pI de respectivement 4 ; 6,9; 6,9 et 9,3 (Figure 54). Pour pouvoir comparer les positions des intermédiaires dans différentes expériences, une mesure de la position d'une forme intermédiaire relativement aux formes réduites et oxydées a été définie. Cette mesure, appelée  $R_f$  est égale à la distance entre un intermédiaire et la forme oxydée divisée par la distance entre la forme réduite et la forme oxydée. Les valeurs de  $R_f$  pour C et D sont respectivement de 0,54 et 0,53.

Le résultat de l'analyse du repliement de Tox62ETag  $^{35}\text{S}$  à 20nM par isoélectrophorèse est présenté en Figure 55. La bande majoritaire au temps 1 heure a un  $R_f$  de 0,51 comparable aux valeurs de l'intermédiaire C et de l'intermédiaire D de Tox62ETag (respectivement 0,53 et 0,54). Il semble donc que cette bande soit composée des intermédiaires C et D de Tox62ETag. Ces intermédiaires ont des pI très semblables c'est pourquoi lorsqu'ils sont tous deux présents, il est difficile de les différencier. Les évolutions de la forme réduite, de cette bande et de la forme oxydée sont comparables à celles obtenues en HPLC (Figure 54).

Ces résultats montrent que pour Tox62ETag le mécanisme et la cinétique de repliement sont conservés que la concentration en protéine soit de 20nM ou de 15 $\mu\text{M}$ .

Cette méthode permet donc de suivre l'évolution des formes intermédiaires pour des concentrations en toxine réduite de l'ordre du nanomolaire.

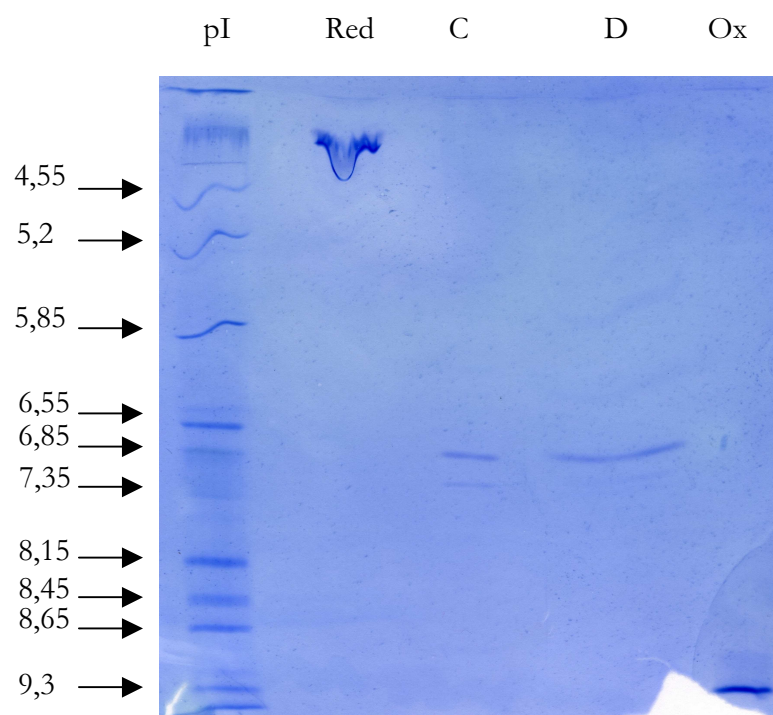


Figure 54 : Gel d'isoélectrophorèse coloré au bleu de Coomassie des intermédiaires C et D de Tox62 Etag purifiés en HPLC et bloqués au MTSBS

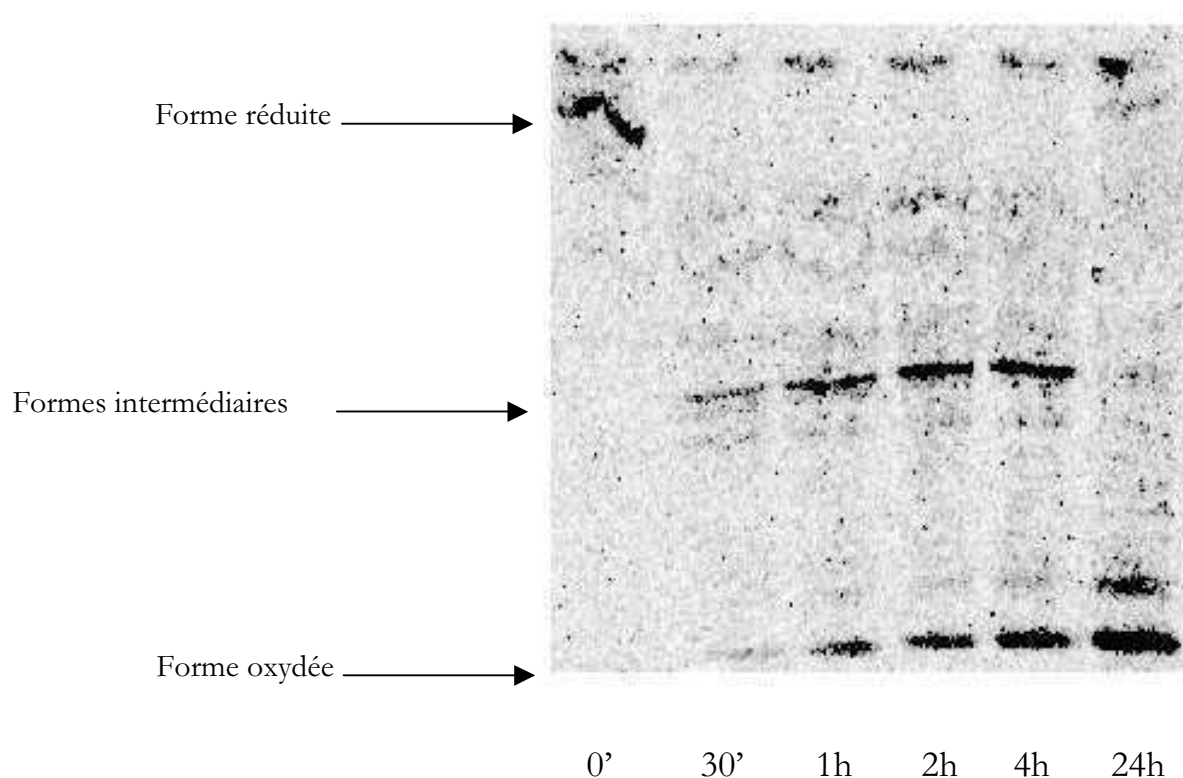


Figure 55 : Repliement de Tox62Etag suivi en isoélectrophorèse à 20 nM

### 2.3.2 Contrôle du repliement par action des anticorps

Dans le but de réaliser la vectorisation du repliement, il est nécessaire que les anticorps utilisés aient un effet sur le repliement de la toxine. En présence du tampon oxydo-réducteur, ces derniers doivent inhiber la formation des ponts dans la partie de la molécule correspondant à leur épitope.

#### 2.3.2.1 Seul acpTx2 inhibe le repliement

- Suivi de l'inhibition du repliement par acpTx2, acpTx3 et acpTxETag par mesure de l'activité

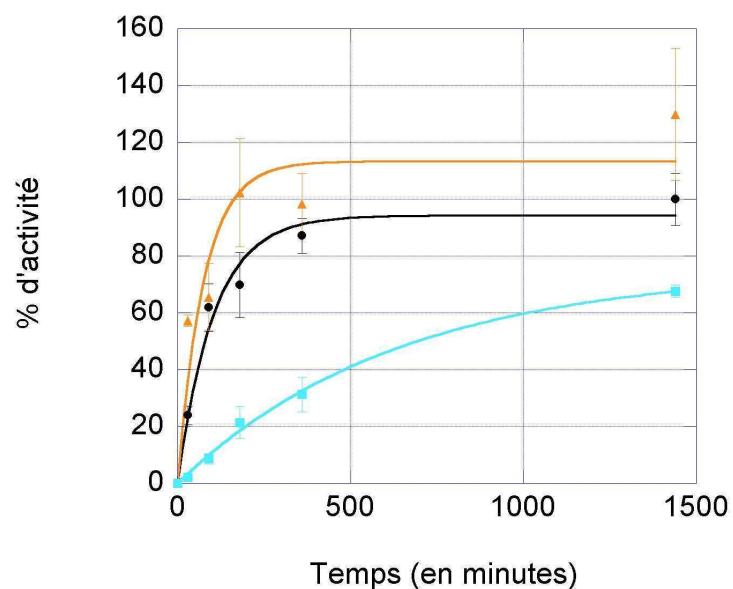
L'action de acpTx2 et de acpTx3 sur le repliement de Tox62 et celle de acpTxETag sur Tox62ETag ont été testées. Les points des courbes présentées en Figure 56 ont été obtenus en calculant un pourcentage d'activité en considérant que la valeur de concentration atteinte après 24 heures de repliement en présence d'immunoglobuline G non spécifiques de lapin (Figure 56 courbes noires) correspond à 100% d'activité.

Après 180 minutes de repliement, on obtient 20% d'activité en présence de l'anticorps anti pTx2 et 70% pour l'expérience contrôle (Figure 56 1). Le repliement de Tox62 est bien inhibé par la présence de cet anticorps. A l'inverse, la cinétique de repliement de Tox62 n'est pas modifiée par la présence de acpTx3.

Une valeur d'activité supérieure à 100% est observée après 24 heures de repliement en présence d'acpTx3. Cette anomalie peut s'expliquer par le fait que la technique utilisée est peu précise. De plus, aux faibles concentrations de toxines utilisées (2nM en l'occurrence), le taux de récupération de toxine est variable. Cette variation a été attribuée à une adsorption de la toxine réduite sur les parois des tubes Eppendorf et des cônes utilisés. La présence de protéines non spécifiques et l'utilisation de cônes et de tubes silanisés (voir matériels et méthodes) ont permis de minimiser ce phénomène sans toutefois l'éliminer totalement. Cependant cette imprécision de mesure ne remet pas en cause l'observation de l'inhibition par acpTx2 car non seulement le taux de récupération est modifié (60%) mais également la cinétique d'apparition de l'activité.

Le taux de récupération de toxine Tox62ETag active en présence de acpTxETag est identique à celui obtenu dans l'expérience témoin (Figure 56 2). La cinétique d'apparition d'activité semble peu affectée par l'action de acpTxETag. Ce résultat n'implique pas nécessairement que cet anticorps ne modifie pas le repliement. En effet, étant spécifique de l'extrémité C-terminale de la toxine, il est possible que cet anticorps soit lié, qu'il empêche la formation du pont C-terminal mais que le reste de la molécule soit structuré. Dans ce cas acpTxETag inhiberait bien la formation du pont mais n'empêcherait pas la formation d'intermédiaires disposant d'une certaine activité biologique. On sait en effet que l'activité biologique de la protéine se situe essentiellement dans le bas des boucles II et III ((Tremeau et al., 1995)), donc assez éloigné de l'épitope. De plus, on a vu que les intermédiaires C et D possèdent une activité biologique. Dans le but de connaître l'effet de acpTxETag sur l'apparition des formes intermédiaires de Tox62ETag, l'étude en isoélectrophorèse semble mieux adaptée.

1



2

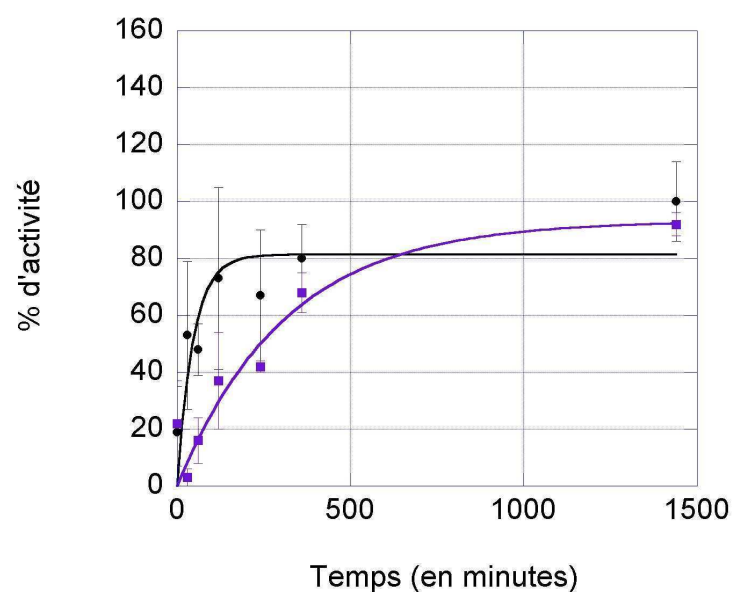


Figure 56 : 1, Repliement de Tox62 à 2nM suivi en activité biologique en présence d'immunoglobuline G non spécifiques de lapin (ronds noirs), de acpTx2 (carrés cyans), de acpTx3 (triangles oranges). 2, Repliement de Tox62 ETag à 2nM suivi en activité biologique en présence d'immunoglobuline G de lapin non spécifiques (ronds noirs) et de acpTxETag (carrés violets)

- Suivi de l'inhibition du repliement par acpTx2, acpTx3 et acpTxETag par isoélectrophorèse

Les images présentées en Figure 57 confirment les résultats obtenus par mesure de l'activité. Il apparaît clairement qu'en présence de acpTx2, le repliement de Tox62ETag <sup>35</sup>S est fortement ralenti. En effet, après 4 heures de repliement, la bande correspondant à la forme oxydée est à peine discernable. La comparaison entre les images A et B de la Figure 57 montre que les espèces à faible nombre de ponts (proche de la forme réduite) restent présentes jusqu'au temps 24 heures ce qui n'est pas le cas dans l'expérience témoin. Outre ces espèces à faible nombre de ponts et la forme native qui apparaît après 24 heures de repliement, une bande ayant un Rf de 0,41 apparaît dès 30 minutes et reste présente même au bout de 24 heures. Comme cela a été décrit en partie 2.3.1.2 de ces résultats cette zone correspond à celle des intermédiaires à 3 ponts (le Rf des intermédiaires à 3 ponts de Tox62ETag est de 0,51). Cette expérience confirme donc l'action inhibitrice de l'anticorps acpTx2 sur le repliement de Tox62ETag. Elle montre également qu'en présence de acpTx2, le repliement de Tox62ETag produit un intermédiaire disposant probablement de trois ponts disulfure mais différent de ceux se formant lorsque cette protéine se replie en l'absence d'anticorps spécifiques.

La même expérience effectuée en présence de acpTx3 montre que cet anticorps n'a d'effet ni sur la vitesse d'apparition de la forme native ni sur la composition en intermédiaires (comparer Figure 57 A et 57 C).

Lorsque le repliement de Tox62ETag est effectué en présence de acpTxETag, la cinétique d'apparition de la forme native est similaire à celle observée dans l'expérience témoin (Figure 57 A et D). Cependant la bande correspondant aux intermédiaires à trois ponts dans l'expérience témoin a dans ce cas un Rf de 0,45 (à comparer à 0,51). La présence de l'anticorps acpTxETag semble impliquer la formation d'un intermédiaire à trois ponts différent de ceux apparaissant lors du repliement sans anticorps.

#### ▪ Conclusions

Les deux méthodes utilisées montrent que seul acpTx2 est capable d'inhiber le repliement de Tox62.

Lorsque cet anticorps est ajouté, un nouvel intermédiaire de repliement apparaît.



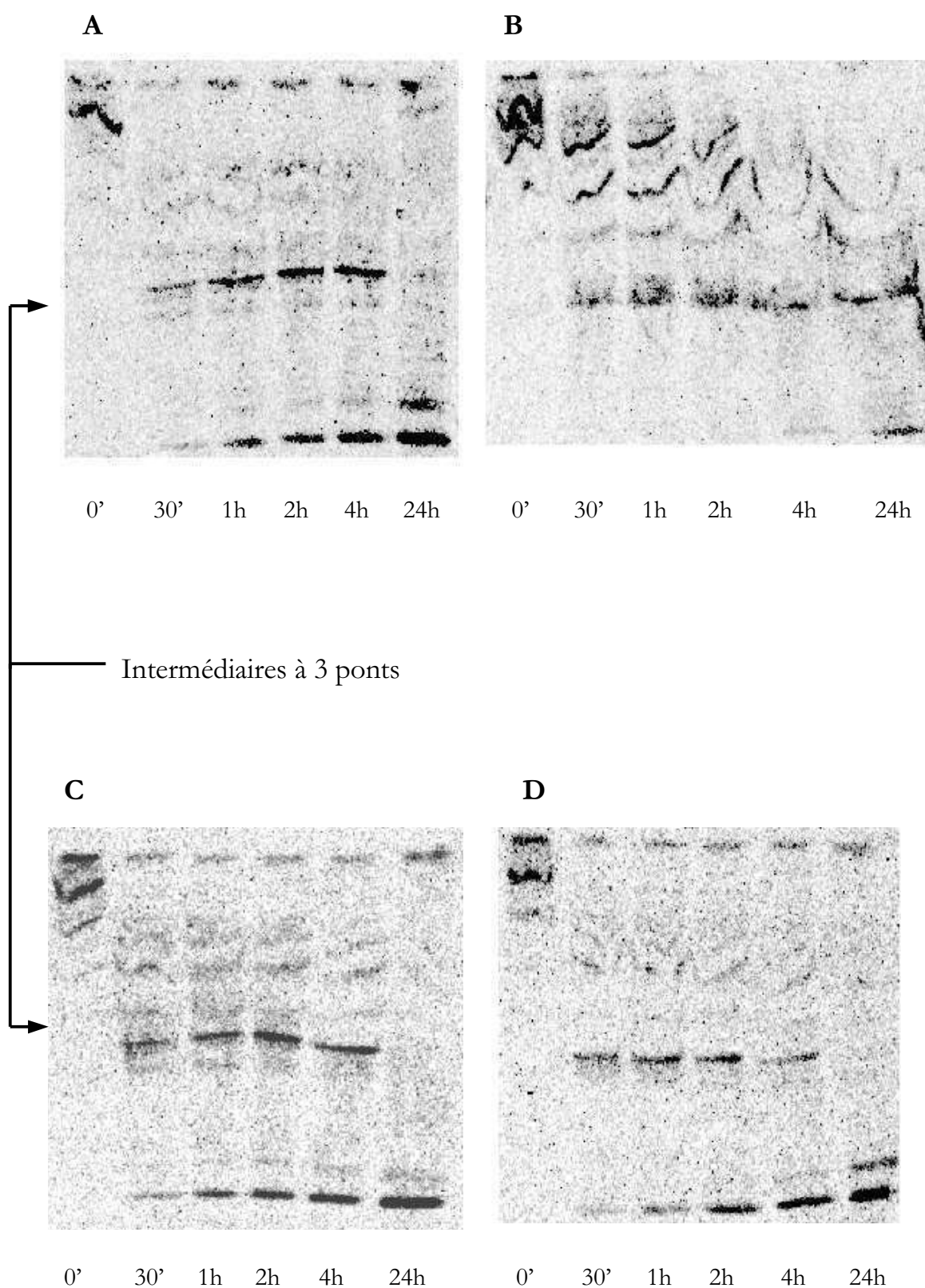


Figure 57 : Repliement de Tox62 ETag suivi en isoélectrophorèse en présence d'immunoglobuline G non spécifiques de lapin (A) et de acpTx2 (B) et de acpTx3 (C) et de acpTxETag (D).

### 2.3.2.2 Levée de l'inhibition de acpTx2 par ajout du peptide épitope

Le seul anticorps ayant un effet sur la cinétique de repliement de Tox62 est acpTx2. L'expérience suivante (Figure 58) permet de vérifier qu'il est possible de lever son inhibition par ajout du peptide épitope pTx2.

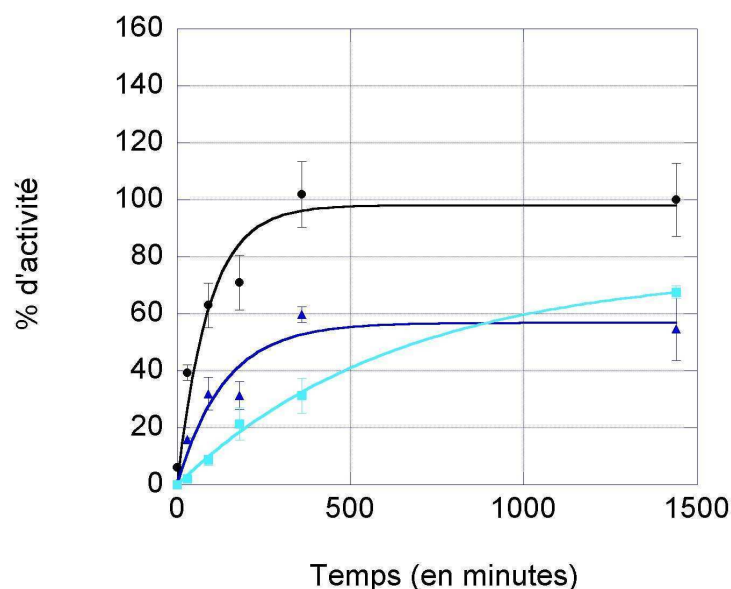


Figure 58 : Repliement de Tox62 à 2nM suivi en activité biologique en présence d'immunoglobuline G de lapin non spécifiques (ronds noirs pleins), de acpTx2 (carrés cyan pleins) et dissociation de acpTx2 par ajout d'un excès de pTx2 au temps zéro (triangles bleus pleins)

L'ajout de peptide épitope au temps 0 permet d'obtenir une cinétique d'apparition de formes actives intermédiaire entre celle de l'expérience témoin et celle de l'expérience d'inhibition par acpTx2. Le taux de récupération après 24 heures dans l'expérience de dissociation est de 70% donc à peine supérieur celui obtenu dans l'expérience d'inhibition (qui est de 60%).

Il semble donc possible (au moins partiellement) de lever l'inhibition du repliement par ajout du peptide épitope.

## 2.4 Conclusions concernant la vectorisation

Les résultats obtenus dans cette étude montrent que la stratégie choisie permet de produire des anticorps reconnaissant la forme réduite de la protéine. La liaison d'un de ces anticorps (acpTx2) permet d'inhiber le repliement de la toxine et ce en présence d'un tampon oxydo-réducteur. Cette inhibition modifie également le mécanisme de repliement de la toxine puisqu'un nouvel intermédiaire à trois ponts apparaît en présence de acpTx2. Il est possible de lever l'inhibition par l'utilisation du peptide épitope. L'anticorps acpTx2 répond donc à l'ensemble des conditions établies dans le cahier des charges. Cependant, pour réaliser une vectorisation du repliement, il serait nécessaire d'obtenir plusieurs anticorps de ce type recouvrant la séquence de la toxine.



## **DISCUSSION**



## 1 LE REPLIEMENT OXYDANT DE LA TOXINE $\alpha$ ET L'INFLUENCE DE LK1

### 1.1 La vitesse de repliement des neurotoxines courtes est dépendante de la longueur de lk1

Les travaux d'Anfinsen dans les années 60 sur le repliement oxydant de la Ribonucléase A ont montré que l'information nécessaire à l'acquisition de la structure tridimensionnelle est contenue dans la séquence en acides aminés ((White, 1961), (Haber and Anfinsen, 1961)). Depuis lors les auteurs qui ont mené des études sur le repliement des protéines *in vitro* ont tenté de découvrir un lien entre la séquence des protéines et les processus impliqués dans leur repliement.

Au début des années 80, la détermination des cinétiques de repliement oxydant de neuf toxines de serpent appartenant à la famille des protéines trois doigts s'inscrivait dans cette démarche((Menez et al., 1980)). Cette étude a permis de mettre en évidence que la vitesse initiale d'apparition des éléments de structure secondaire dans ces protéines est corrélée au nombre de résidus dans leurs séquences et plus particulièrement à la longueur d'une boucle reliant les deux premiers doigts appelée dans la présente étude lk1. A l'issue de cette étude, il n'était pas possible d'affirmer que les différences de cinétique étaient uniquement dues à la longueur de cette boucle car il existait d'autres modifications dans la séquence de ces 9 protéines.

Dans le but de tester l'influence de la longueur de cette boucle sur la vitesse de repliement toutes choses égales par ailleurs, Ruoppolo *et al.* ont produit des mutants de la toxine  $\alpha$  avec différentes longueurs de lk1((Ruoppolo et al., 1998)). Ces mutants correspondent aux mutants Tox60, Tox61 et Tox62 de la présente étude. Leurs résultats ont confirmé la relation selon laquelle la longueur de lk1 influe sur la cinétique de repliement. Pour expliquer le lien entre l'addition d'un résidu dans la séquence de lk1 et la diminution de la vitesse de repliement, ces auteurs ont avancé l'hypothèse selon laquelle un lk1 plus long impliquerait l'existence d'un espace conformationnel plus vaste dans cette région de la molécule et donc dans les étapes précoces du repliement, un nombre d'intermédiaires plus important qui constitueraient des pièges cinétiques ralentissant le phénomène global.

### 1.2 Rejet de l'hypothèse d'une augmentation du nombre d'intermédiaires

Les chromatogrammes d'analyse des mélanges de repliement aux temps courts de Tox60, Tox61 et Tox62, montrent qu'il n'existe pas de différence dans le nombre d'intermédiaires apparaissant dans les étapes précoces et qu'au contraire ces profils sont semblables. L'intégration de ces chromatogrammes associée à la détermination par analyse de masse du nombre de ponts dans chacune des espèces isolées ont permis de détailler l'évolution de groupes d'espèces : réduite, à un pont, à deux ponts et à trois ou quatre ponts (il n'est pas possible de séparer les espèces à trois ponts des espèces à quatre ponts par la méthode de chromatographie en échange d'ions utilisée). Il s'avère que lorsqu'on compare leur évolution pour Tox60 et Tox61, les formes à 1 et 2 ponts disparaissent plus vite pour générer les formes à trois et quatre ponts dans le cas de

Tox60 que dans le cas de Tox61. Une partie de l'augmentation de vitesse de Tox60 est donc liée au fait que les étapes précoces sont plus « efficaces ».

Cependant cette différence ne semble pas impliquer de modification dans la nature des intermédiaires formés mais plutôt dans leur comportement cinétique. Lorsqu'on compare l'évolution de ces mêmes groupes d'espèces entre Tox61 et Tox62, il n'y a aucune différence. Le groupe des espèces à trois et quatre ponts apparaît à la même vitesse dans les deux cas. La différence de cinétique d'apparition de la forme native observée entre Tox61 et Tox62 est donc uniquement due à une différence d'efficacité de conversion des formes à trois ponts et quatre ponts en forme native.

Les résultats de cette série d'expériences ont montré que les différences de cinétique de repliement ne sont pas dues à une modification dans le nombre d'intermédiaires apparaissant dans les étapes précoces de repliement et ne sont donc pas compatibles avec l'hypothèse proposée.

### **1.3 Le comportement cinétique de deux formes à trois ponts détermine l'efficacité globale**

Pour obtenir de plus amples informations sur l'origine des différences de cinétique de repliement, les étapes finales ont été analysées par HPLC en phase inverse. L'analyse de l'évolution des profils chromatographiques montre que dans les trois cas (Tox60, Tox61 et Tox62) il se forme deux intermédiaires à trois ponts. Leur accumulation détermine la vitesse globale de repliement.

Dans le but d'analyser plus précisément les différences de cinétique de repliement, l'étude s'est ensuite focalisée sur la comparaison entre le repliement de Tox61 et celui de Tox62. Les intermédiaires à trois ponts de Tox61 ont été caractérisés par digestion à la trypsine et analyse de masse puis éventuellement séquençage des peptides obtenus. Cette analyse a permis de montrer que l'intermédiaire D de Tox61 est un intermédiaire à 3 ponts dont les cystéines 17 et 40 sont libres tout comme ce qui avait été montré par Ruoppolo *et al.* pour Tox62 ((Ruoppolo et al., 2001)). Concernant l'intermédiaire C de Tox61, tous les résultats obtenus en analyse de masse sont compatibles avec l'hypothèse selon laquelle les cystéines libres de cet intermédiaire sont les cystéines 42 et 53. La comparaison du spectre de dichroïsme circulaire de cette espèce avec celui de Tox61pT3 indique que leurs structures secondaires sont identiques.

Ces éléments nous ont conduit à admettre pour le reste de l'étude que les intermédiaires C et D de Tox61 sont les intermédiaires des[43-54] et des[17-41], respectivement. La différence de vitesse de repliement entre Tox61 et Tox62 n'est donc pas liée à une modification de la nature des espèces à trois ponts qui apparaissent mais à une modification de leur comportement cinétique.

## 1.4 Un basculement de voie productive modifie la cinétique

Ce comportement a été étudié plus en détail en analysant le repliement en condition uniquement oxydante de Tox61 et Tox62 en partant de leurs intermédiaires C et D purifiés. En analysant le repliement en présence de glutathion oxydé, l'objectif est de simplifier l'interprétation des résultats. En effet, dans ces conditions, l'influence d'une éventuelle réduction de certaines espèces est supprimée. Un réarrangement intramoléculaire (reshuffling) reste toutefois possible. Deux différences majeures ont pu être identifiées : l'unique voie productive du repliement de Tox61 passe par l'intermédiaire D alors que dans le cas de Tox62 cette voie est moins efficace, la voie sans issue que constitue l'intermédiaire C dans le repliement de Tox61 est productive dans Tox62 (Figure 59).

La simple addition d'un résidu dans le lk1 de la toxine  $\alpha$  a impliqué un basculement de la voie productive, voie D dans le cas de Tox61 et C dans le cas de Tox62.

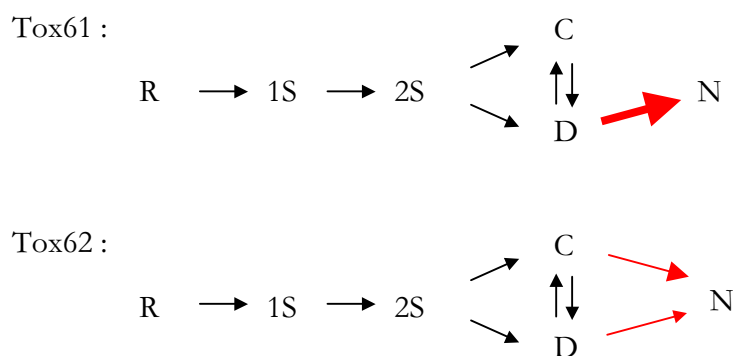


Figure 59 : Mécanismes cinétiques et voies productives de repliement de Tox61 et Tox62.

Une analyse de la bibliographie sur le repliement des protéines à ponts disulfure montre que c'est la première fois que ce type de résultat a été obtenu. Un résultat approchant a été obtenu par Chang *et al.* dans le cadre de l'étude du repliement oxydant de l' $\alpha$ -lactalbumine ((Chang and Li, 2002)). En l'absence de calcium, le repliement oxydant de l' $\alpha$ -lactalbumine produit un grand nombre d'intermédiaires à 1, 2 et 3 ponts et des intermédiaires à 4 ponts mésappariés. Au contraire, en présence de calcium, le repliement de cette même protéine produit 2 intermédiaires prédominants contenant uniquement des ponts natifs et notamment deux ponts situés dans la partie de la molécule fixant le calcium. La liaison du calcium à la molécule semble constituer dans ce cas un point de nucléation pour le repliement de cette protéine. La simple addition de calcium dans le milieu de repliement implique un basculement de mode de repliement d'un mode de type BPTI à un mode de type hirudine ((Arolas et al., 2006)). Le basculement observé dans le cas de l' $\alpha$ -lactalbumine concerne les étapes précoces du repliement alors que dans notre cas ce sont les étapes finales du repliement qui sont modifiées.



## 1.5 Explication structurale de la différence de productivité de la voie C

Nous avons cherché à proposer une explication structurale à ce phénomène en résolvant la structure tridimensionnelle d'analogues des intermédiaires à trois ponts de ces deux mutants. La comparaison des structures des analogues structuraux des intermédiaires C (Tox61pT3 et Tox62pT3) par RMN a permis d'identifier leurs différences structurales. Dans le cas de Tox61pT3 cette étude a mis en évidence l'existence d'un équilibre entre deux conformations à une échelle de temps comprise entre 10 et 100ms. La première de ces formes (forme A) a une structure identique à celle de la molécule native. La deuxième forme (forme B) a toute sa partie C-terminale déstructurée, celle-là même qui est maintenue par le pont 42-53 dans la molécule native, (Figure 60). En revanche, le variant Tox62pT3 n'existe que sous une seule forme structurale identique à la structure native. La comparaison des structures cristallographiques de Tox61 ((Gilquin et al., 2003)) et de Tox62 (résolue par Enrico Stura dans le cadre de la présente étude) montre que seuls les atomes du lk1 sont déplacés, le reste de la structure étant parfaitement identique. Les résultats de RMN permettent de considérer que les structures des intermédiaires C (forme A pour Tox61pT3) sont très proches des structures des protéines natives. Les structures cristallines constituent donc de bons modèles de ces intermédiaires et peuvent donc être utilisées pour comparer leurs structures. Le rappel effectué en introduction concernant les études sur la structure d'analogues d'intermédiaires de repliement auxquels il ne manque qu'un pont, a montré que ces analogues ont des structures très similaires à celle de la protéine native ((Eigenbrot et al., 1990), (van Mierlo et al., 1991), (Shimotakahara et al., 1997), (Laity et al., 1997), (van den Berg et al., 1999), (Inaka et al., 1991), (Arolas et al., 2005)). Les cas de la toxine  $\alpha$  et du mutant Tox62 ne sont donc pas exceptionnels.

Les trois raisons pouvant expliquer l'incapacité d'un intermédiaire à former le dernier pont manquant ont été présentées en introduction. Il peut s'agir (i) d'un enfouissement des cystéines libres ((Narayan et al., 2000)). Dans le cas où les cystéines libres s'avéreraient accessibles, la difficulté à fermer le pont peut provenir soit (ii) de la nécessité de passer par un état structural contraint ((Gilbert, 1990)) ou (iii) d'une entropie conformationnelle importante dans la région du pont ((Gilbert, 1990)). Les structures cristallographiques de Tox61 et Tox62 montrent que les cystéines du pont 3 de chacune de ces molécules sont accessibles au solvant. L'impossibilité de fermeture de ce pont dans l'intermédiaire C de Tox61 ainsi que la faible efficacité de conversion de l'intermédiaire C de Tox62 ne peuvent être expliquées par l'hypothèse (i). Sachant que l'intermédiaire C de Tox62 a une structure native, l'hypothèse (ii) est la seule restante. La faible vitesse de production de la forme native à partir de l'intermédiaire C de Tox62 semble due à l'existence de contraintes géométriques défavorables : pour que la cystéine libre se trouve à proximité du disulfure mixte, il est probable que la molécule doive passer par un état structural contraint, impliquant une faible concentration locale en ions thiolates à proximité du disulfure mixte.

Pour le cas de l'intermédiaire C de Tox61, la situation est différente. Les structures de Tox61pT3 forme A et Tox62pT3 étant identiques à leur molécule native respective et sachant que Tox61 et Tox62 ont des structures identiques, on peut en déduire qu'il n'existe pas de différence importante de structure entre elles. L'unique différence structurale significative révélée par cette étude en RMN concerne l'existence d'un échange conformationnel entre deux formes dans le cas de Tox61pT3. Cet échange est associé à une diminution de la stabilité thermique se traduisant par une différence de température de fusion de 9°C entre Tox61pT3 et Tox62pT3.

1



2



Figure 60 : Structure cristallographique de la toxine  $\alpha$  de *Naja nigricollis* (1IQ9). 1 :vue de devant, 2 : vue de derrière. En magenta les résidus dont les valeurs de nOe hétéronucléaire sont supérieures à 0,2. La proline 47 est également figurée en magenta même si elle ne dispose pas de proton amide. Les atomes de soufre du pont disulfure 42-53 ont été supprimés.

Les données structurales obtenues suggèrent que les différences de productivité des intermédiaires C de Tox61 et Tox62 sont dues à l'existence ou non de cet échange conformationnel. Lorsque l'intermédiaire C de Tox61 est dans la conformation B, il est probable que les cystéines 42 et 53 soient très éloignées l'une de l'autre (Figure 60) et donc que la fermeture du pont correspondant soit impossible. L'intermédiaire C de Tox62 qui a une structure identique à la forme A de Tox61 est productif même si sa vitesse de fermeture est faible comparée à celle de la fermeture du pont D de Tox61. En conséquence, la forme A de Tox61pT3 devrait être productive si ce n'était l'échange permanent avec la forme B. L'impossibilité de fermeture du pont 42-53 dans le cas de Tox61 semble indiquer qu'il est nécessaire que la molécule soit maintenue un certain temps dans la conformation A pour que le pont puisse se former. Ce temps doit certainement être supérieur à 10 ms.

## 1.6 Origine de la stabilisation de la forme structurée de l'intermédiaire C

Cette explication structurale de la différence de comportement cinétique des intermédiaires C de Tox61 et Tox62 maintenant établie, il reste à tenter de comprendre comment l'addition d'un résidu dans lk1 stabilise la forme structurée et empêche l'apparition de la forme B de l'intermédiaire C.

La partie 1.3.4 des résultats a permis de démontrer que les structures des molécules natives sont de bons modèles des structures de Tox61pT3 forme A et de Tox62pT3. La comparaison des structures cristallographiques des toxines natives montre que les seuls résidus dont la position spatiale est modifiée par l'addition d'une sérine dans lk1 sont justement localisés dans la boucle lk1 (Figure 36 en partie 1.3.1 des résultats).

Deux hypothèses (qui ne s'excluent pas mutuellement) peuvent être proposées pour expliquer l'origine de la stabilisation de la forme A de l'intermédiaire C lorsqu'une sérine est ajoutée dans lk1 :

- Hypothèse liée au nombre de résidus dans la boucle lk1: la structure native de la toxine  $\alpha$  de *Naja nigricollis* est peut-être « frustrée » (dans le sens de minimum de frustration proposé dans l'article (Bryngelson and Wolynes, 1987)) par le fait que le linker 1 est court (5 résidus). La présence de 6 résidus comme c'est le cas dans Tox62 pourrait diminuer les tensions existant au sein de la molécule et stabiliser la forme native, y compris dans l'intermédiaire C.

- Hypothèse liée à l'influence du positionnement du glutamate 20 (appelé dans ce paragraphe E20 pour Tox61 et E21 pour Tox62) : une analyse détaillée des structures cristallographiques de Tox61 et Tox62 dans la zone de lk1 fait ressortir que E20 (E21 pour Tox62) est l'unique résidu qui, à la fois voit sa position modifiée par la mutation et interagit avec d'autres régions de la molécule (Figure 61). La fonction acide de l'extrémité de la chaîne latérale de E20 est orientée vers le solvant dans le cas de Tox61 et vers l'intérieur de la structure dans le cas de Tox62. Dans la structure de Tox62 un atome d'oxygène de cette fonction acide forme une liaison hydrogène (distance de 3,5 Å) avec le proton amide de la chaîne principale de l'asparagine 23. Ces observations sont corroborées par les informations obtenues en RMN. En effet, les contacts nOe du résidu E21 de Tox62 (présentés en annexe 5) indiquent que son proton  $\gamma$  établit des contacts avec le proton amide de la chaîne principale de l'asparagine 23, avec le proton amide et le proton  $\alpha$  de la glycine 42 et avec le proton amide de la cystéine 43. Aucun de ces contacts n'est observé pour le proton  $\gamma$  du résidu E20 de Tox61 (annexe 5). L'ensemble de ces résultats permet de proposer que l'addition d'un résidu dans lk1 implique une modification de l'orientation de la chaîne latérale du résidu E21. Il pointe alors vers l'intérieur de la molécule, forme une liaison hydrogène et éventuellement d'autres contacts conduisant à une moindre mobilité dans la partie apicale du troisième doigt. Cette modification semble au moins en partie responsable de la stabilisation de la forme repliée de l'intermédiaire C.

## **1.7 Interprétation structurale de la perte d'efficacité de la voie D**

Les résultats précédents expliquent en partie le basculement de voie productive par l'ouverture de la voie C lors de l'addition d'un résidu dans lk1. Cependant, ils n'expliquent pas l'importante diminution de vitesse de repliement due à une moindre efficacité de la voie D dans le cas du repliement de Tox62.

L'étude en RMN des analogues structuraux des intermédiaires D de Tox61 et Tox62 est en cours. Ces résultats devraient permettre de proposer une explication au ralentissement de la voie D lorsqu'un résidu est ajouté dans lk1, ce qui conduit à un repliement global plus lent.

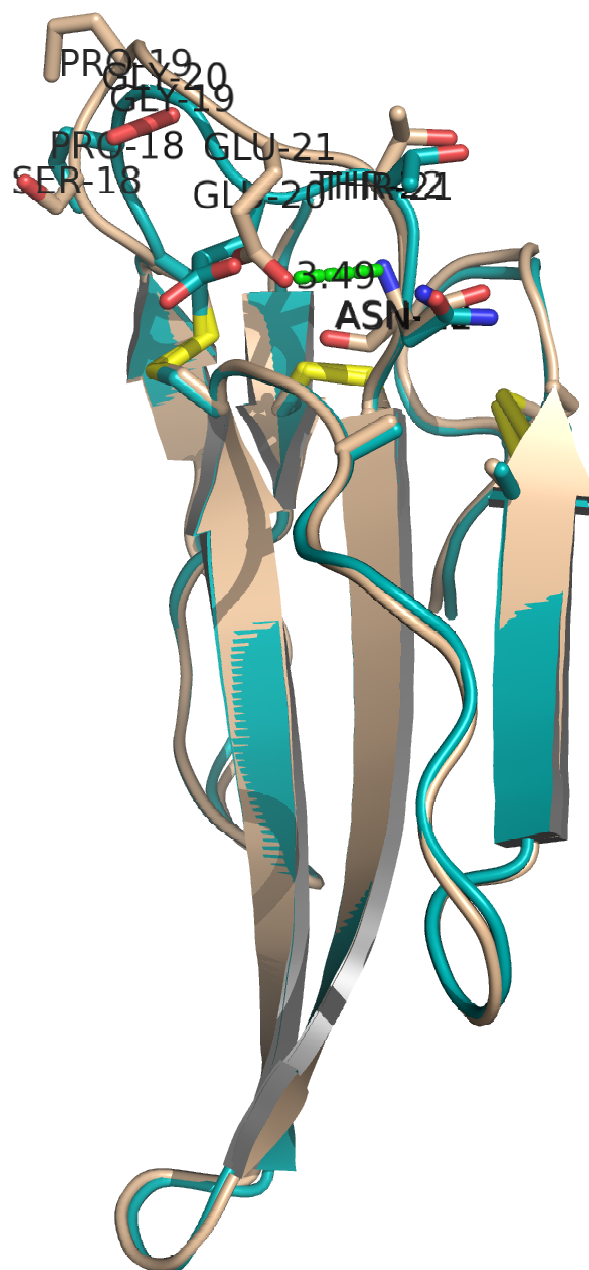


Figure 61 : Superposition des structures cristallographiques de la toxine  $\alpha$  de *Naja nigricollis* (1IQ9, en bleu) et de la structure de Tox62 (en beige). Vue de la tranche. Les atomes des chaînes latérales des résidus de Ik1 sont présentés en bâtonnets. La distance entre l'atome d'oxygène de la chaîne latérale de E21 et le proton amide de N23 dans Tox62 est indiquée en vert. Les atomes de soufre du pont disulfure 42-53 ont été supprimés.

## 1.8 Généralisation des résultats à la famille des neurotoxines courtes

La toxine  $\alpha$  de *Naja nigricollis* se replie donc en formant deux intermédiaires à trois ponts natifs des[42-53] et des[17-40]. La fermeture du pont 17-40 constitue l'étape finale et limitante du repliement de cette molécule. L'addition d'une sérine dans lk1 induit une diminution de la vitesse de repliement sans que la nature des intermédiaires soit modifiée. Ménez *et al.* ((Ménez et al., 1980)) avaient émis l'hypothèse que les différences de vitesse de repliement d'un jeu de 9 neurotoxines courtes provenaient de la différence du nombre de résidu dans lk1. Pour pouvoir généraliser le résultat obtenu dans la présente étude, il serait nécessaire de vérifier au préalable que ces différentes neurotoxines courtes se replient en formant deux intermédiaires à trois ponts identiques à ceux observés lors du repliement de la toxine  $\alpha$ . La question reviendrait alors à vérifier que différentes protéines appartenant à la même famille structurale partagent un mécanisme de repliement commun.

Cette question a été analysée en détail dans le cas des protéines ne disposant pas de ponts disulfure ((Zarrine-Afsar et al., 2005)). Dans cette revue, les états de transition de 7 paires de protéines appartenant à la même famille structurale ont été caractérisés par la méthode des  $\phi$ -values puis comparés. Les auteurs ont conclu de cette analyse que dans certaines familles l'état de transition est conservé alors que dans d'autres il ne l'est pas, ces résultats ne permettant donc pas de trancher la question.

Dans le domaine du repliement des protéines riches en ponts disulfure, ce type d'analyse globale n'est pas envisageable car d'une part le nombre de protéines dont le mécanisme de repliement a été analysé en détail reste faible (voir Tableau 1) et pour la plupart de ces études une seule protéine de la famille structurale à laquelle appartient le modèle a été étudiée. Il n'existe qu'un seul article ayant pour objet la comparaison des mécanismes de repliement oxydant de deux protéines partageant le même fold, il concerne le BPTI et TAP ((Chang, 1996), (Chang and Li, 2005)). Les conclusions de cet article indiquent que les mécanismes de repliement de ces deux molécules divergent fortement.

Il faut noter que le mécanisme de repliement d'une autre toxine trois doigts, la cardiotoxine III, a été étudié ((Chang et al., 2006)). Le repliement oxydant de cette toxine produit une grande diversité d'intermédiaires et une quantité importante de formes à 4 ponts mésappariés. La classification établie par Arolas *et al.* apparente ce type de repliement à celui de l'hirudine ((Arolas et al., 2006)). Les résultats de notre étude montrent que la toxine  $\alpha$  produit un nombre restreint d'intermédiaires et qu'aucune forme à 4 ponts mésappariés n'apparaît. Ce type de repliement peut être rapproché de celui du BPTI. La comparaison des mécanismes de repliement de ces deux toxines trois doigts indique qu'ils sont radicalement différents (voir article présenté en annexe 6).

L'analyse des mécanismes de repliement des différentes protéines étudiées par Ménez *et al.* dans les années 80 constituerait une perspective intéressante de prolongement de cette étude car elle permettrait d'apporter des éléments de réponse supplémentaires à la question de la similitude de la voie de repliement de protéine proches d'un point de vue structural dans le cas des protéines à ponts disulfure. Dans l'hypothèse où effectivement, le mécanisme de repliement serait conservé au sein des neurotoxines courtes, il serait envisageable de généraliser l'influence de la longueur d'une boucle sur le mécanisme de repliement d'une famille structurale de protéines.

## 2 VECTORISATION *IN VITRO* DU REPLIEMENT DE LA TOXINE $\alpha$

### 2.1 Influence de la vectorisation sur le repliement

L'étude menée dans la partie précédente s'inscrit dans une méthodologie habituellement utilisée pour décrire le mécanisme de repliement des protéines à ponts disulfure *in vitro* ((Arolas et al., 2006)).

Lorsqu'une protéine est synthétisée au sein d'une cellule vivante, il est possible que les phénomènes impliqués dans le repliement de cette protéine soient différents de ceux décrits par ces expériences réalisées *in vitro*. Plusieurs études du repliement *in vivo* de certaines protéines ont mis en évidence que des éléments de structure peuvent apparaître dans les chaînes polypeptidiques en cours d'élongation : des ponts disulfure ((Bergman and Kuehl, 1979), (Chen et al., 1995)), un épitope ((Fedorov et al., 1992), (Friguet et al., 1993), (Tokatlidis et al., 1995)), un domaine catalytique ((Nicola et al., 1999)). Ces résultats ne constituent pas une démonstration de la généralité du repliement cotraductionnel, et savoir si les protéines se replient de manière co ou post traductionnelle demeure une question ouverte ((Basharov, 2003)).

Cependant, si l'hypothèse selon laquelle les protéines se replient systématiquement *in vivo* de façon cotraductionnelle est admise, il est légitime de se demander si le mécanisme de repliement tel qu'il est étudié *in vitro* est représentatif des processus qui ont lieu dans la cellule. En effet, *in vitro*, l'expérience commence par une dénaturation totale de la protéine qui est ensuite placée dans un milieu permettant son repliement. Le repliement peut commencer dans n'importe quelle partie de la molécule. A l'inverse, dans le cas du repliement cotraductionnel, la partie N terminale de la protéine apparaît en premier et commence à se replier alors que le reste de la molécule n'est pas synthétisé. Cette partie N-terminale va donc avoir la possibilité de se replier sans être influencée par des interactions avec des résidus situés en aval dans la séquence. Ce type de repliement a été baptisé repliement vectoriel ((Alexandrov, 1993)).

Plusieurs équipes ont tenté d'évaluer l'influence de l'aspect vectoriel du repliement sur la trajectoire de repliement en réalisant la modélisation de la même molécule de façon vectorielle ou non. Les conclusions obtenues divergent. Certaines modélisations concluent que la vectorisation du repliement produit des trajectoires identiques alors que d'autres aboutissent au résultat inverse ((Morrissey et al., 2004), (Huard et al., 2006), (Wang and Klimov, 2008)).

A ce jour, aucun exemple d'étude expérimentale de l'influence de la vectorisation sur le repliement n'est apparu dans la littérature. Seul un article récent de Svetlov *et al.* s'approche de cette question ((Svetlov et al., 2006)). Ces auteurs ont montré que lorsque la luciférase est produite par traduction *in vitro* le rendement de repliement de cette protéine est nettement amélioré. De plus leurs résultats indiquent que ce gain de rendement ne nécessite pas l'action de protéines chaperons de la famille HSP70 (voir partie 2.2.2.4 de l'introduction). A moins qu'il existe d'autres facteurs non identifiés améliorant le rendement de repliement dans ce système de traduction *in vitro*, cette expérience montre que la vectorisation du repliement peut améliorer le rendement d'une protéine composée de deux domaines.

## 2.2 Principe de la vectorisation du repliement par des anticorps

Dans la deuxième partie de mon travail de thèse l'objectif était de concevoir une expérience permettant de recréer artificiellement la vectorisation du repliement d'une protéine *in vitro*. Ce procédé doit permettre d'obtenir « plus aisément » des informations sur l'influence de la vectorisation sur le mécanisme de repliement que lorsque cette protéine est noyée dans le milieu intracellulaire. Il doit permettre d'éliminer l'influence des protéines chaperons et d'éviter d'avoir à isoler la protéine étudiée.

La stratégie générale adoptée est présentée en partie 7.2 de l'introduction et avec la figure 19. En résumé, le but est de dénaturer et réduire une protéine à ponts disulfure, puis d'ajouter des anticorps dirigés contre certaines parties de cette protéine dans sa forme réduite. La liaison de ces anticorps à la protéine réduite doit l'empêcher de se replier et ce même lorsque l'agent oxydant est ajouté. En présence de l'agent oxydant l'inhibition du repliement des anticorps serait levée de manière récurrente en partant de l'extrémité N-terminale pour finir à l'extrémité C-terminale. Cette levée de l'inhibition serait réalisée par l'ajout d'un large excès de peptide épitope.

Dans le cas présent cette méthodologie serait appliquée à une petite protéine riche en ponts disulfure : la toxine  $\alpha$  de *Naja nigricollis*. Plus précisément c'est le mutant Tox62 qui sera utilisé. En effet, l'objectif ultime est d'observer une augmentation de la vitesse de repliement lorsqu'il est vectorisé. La toxine Tox62, dont le repliement est peu rapide, paraît donc constituer un bon modèle d'étude.

## 2.3 Les anticorps obtenus

Des peptides épitopes ont été conçus afin d'obtenir des anticorps spécifiques de 4 portions non chevauchantes de la forme réduite de Tox62. Nous avons considéré que la réussite du reste de l'étude serait probablement déterminée par le choix des séquences de ces épitopes. Ce choix a été réalisé selon les critères suivants. Nous avons utilisé des portions de séquence de la protéine Tox62 d'une longueur comprise entre 10 et 15 résidus pour que cette longueur soit légèrement supérieure à celle habituellement observée pour les antigènes continus qui est de 6 à 10 résidus ((Van Regenmortel, 2001)). Les limites de ces peptides ont été choisies de façon à ne pas recouvrir des éléments de structure secondaire dans leur entier (par exemple il fallait éviter de choisir un peptide recouvrant la séquence du  $\beta$ -hairpin du doigt 2 de Tox62), car ces séquences auraient pu acquérir par elle-même une structure proche de celle de la protéine oxydée. Pour la même raison, nous avons évité de choisir des séquences contenant deux cystéines du même pont. Enfin, les cystéines ont été remplacées par des sérines pour éviter la formation de multimères.

Ces peptides nous ont servi à immuniser des souris pour obtenir des anticorps monoclonaux et des lapins pour obtenir des sérums polyclonaux. Malheureusement, nous n'avons obtenu aucun anticorps monoclonal. L'ensemble de l'étude a donc été effectué avec des solutions polyclonales. Une solution polyclonale est constituée d'une multitude d'espèces d'anticorps disposant d'affinités variables, et la proportion d'immunoglobulines de classe G spécifiques de l'antigène utilisé pour l'immunisation est inférieure à 1% (nous avons obtenu des valeurs de 0,6% pour acpTx2 et 0,2% pour acpTx3). Cette faible proportion implique que les concentrations en anticorps spécifiques sont très faibles (de l'ordre de la centaine de nanomolaire) et que ces solutions sont polluées par une grande quantité d'autres protéines (en l'occurrence essentiellement des immunoglobulines G) qui ne possèdent pas l'activité souhaitée.



Nous avons donc été contraints de développer des méthodes permettant de suivre la cinétique de repliement de Tox62 à une concentration de l'ordre du nanomolaire ce qui s'est avéré très fastidieux. De surcroît, l'utilisation d'anticorps polyclonaux constitue une limitation pour l'interprétation des résultats car l'action de ces mélanges d'anticorps sur le repliement est nécessairement inhomogène. Si cette étude devait être prolongée ou utilisée pour étudier le repliement d'autres protéines, il serait indispensable d'obtenir des anticorps monoclonaux. L'étude décrite ci-après, effectuée avec des solutions polyclonales, a toutefois permis d'évaluer la faisabilité de l'utilisation d'anticorps pour « vectoriser » le repliement. Trois anticorps polyclonaux ont été obtenus et caractérisés : acpTx2, acpTx3 et acpTxETag.

## **2.4 Un seul anticorps répond au cahier des charges**

Pour qu'un anticorps puisse être utilisé pour vectoriser le repliement il doit remplir le cahier des charges suivant :

1. Être spécifiques d'une partie de la séquence de la protéine dans son état réduit,
2. La vitesse de dissociation de ces anticorps doit être rapide par rapport à la vitesse de repliement de Tox62 pour que le démarrage du repliement dans la partie concernée soit synchronisé dans toute la population de protéines étudiées.
3. Il nous faut montrer qu'il est possible de lier plusieurs de ces anticorps sur la même protéine réduite
4. Une fois lié, il doit être capable d'empêcher la fermeture des ponts et donc d'inhiber le repliement en présence d'un agent oxydant,

### **2.4.1 Reconnaissance de la forme réduite**

La première vérification a consisté à comparer l'affinité de nos anticorps pour leur peptide épitope à leur affinité pour la toxine polyS, donc dans le contexte de la protéine entière. Pour les trois anticorps étudiés ces affinités sont comparables. Seul l'anticorps acpTx2 a une affinité diminuée d'un ordre de grandeur lorsque l'épitope est inclus dans la protéine entière. Cette différence peut s'expliquer par une différence de présentation de l'antigène entre la protéine entière et le peptide seul.

Dans un deuxième temps, l'affinité de ces anticorps pour la protéine réduite a été comparée à leur affinité pour la toxine polyS. Ces antigènes diffèrent uniquement par le remplacement des sérines par des cystéines libres. La chaîne latérale des sérines porte une fonction alcool alors que celle des cystéines réduites une fonction thiol. Les fonctions thiol et alcool diffèrent de part la différence d'électronégativité des atomes de soufre et d'oxygène et de part leur différence de pKa (la fonction thiol est beaucoup plus acide que la fonction alcool : leurs pKa sont de 11 et 15 respectivement). Les anticorps acpTx2 et acpTxETag ne sont pas influencés par le remplacement des sérines par des cystéines réduites. L'affinité de acpTx3 est diminuée de deux ordres de grandeur dans le cas de la toxine réduite par rapport à la toxine polyS. Le peptide épitope pTx3 contient deux cystéines mutées en sérine en position centrale de l'épitope ce qui semble ne pas être sans conséquence sur l'affinité de cet anticorps.

Ces résultats montrent que la stratégie adoptée pour obtenir des anticorps reconnaissant la protéine réduite est valable.

Cependant, la substitution des cystéines par des sérines a un effet sur l'affinité de ces anticorps, il est donc préférable d'éviter de choisir des épitopes contenant plusieurs cystéines en position centrale. Nous avons vérifié que l'affinité de nos anticorps pour la forme réduite était plus importante que leur affinité pour la forme oxydée. Le gain d'affinité est de 500 fois pour acpTx2 contre 10 fois pour acpTx3 et acpTxETag.

La stratégie envisagée permet donc d'obtenir des anticorps dirigés contre différentes parties de la toxine et spécifiques de sa forme réduite par rapport à sa forme oxydée.

## **2.4.2 Vitesse de dissociation**

L'étude de la vitesse de dissociation en RIA de l'anticorps acpTx2 a montré que le temps de demi-dissociation est de l'ordre de 3,5 minutes ce qui est 23 fois inférieur au temps de demi-repliement de Tox62 de 4 heures. Cette différence semble suffisamment importante pour considérer comme quasi-instantanée la dissociation de la toxine.

## **2.4.3 Complexe tripartite**

Pour que la vectorisation du repliement d'une protéine par l'utilisation d'anticorps soit envisageable, il faut pouvoir lier au moins deux anticorps sur le même antigène.

Cette question a été étudiée par Biacore dans le cas le plus défavorable, celui de la liaison simultanée de acpTx2 et acpTx3 dont les séquences épitopes ne sont séparées que par trois acides aminés. Les résultats obtenus en Biacore indiquent qu'il existe dans les solutions polyclonales de acpTx2 et acpTx3, des populations d'anticorps capables de se lier simultanément sur le même antigène. Cependant, il subsiste un doute quant à la validité de cette interprétation.

Pour obtenir une image plus directe de la liaison tripartite, il est préférable de tenter d'immobiliser un premier anticorps sur une puce de Biacore puis de tenter de lier un complexe préformé du deuxième anticorps avec l'antigène. Des essais ont été menés pour réaliser ce type d'expérience. Malheureusement, lorsqu'on tente d'immobiliser un des anticorps, comme il s'agit d'une solution polyclonale de faible concentration, il est nécessaire de lier simultanément une grande quantité d'anticorps non spécifiques. Ces molécules inactives produisent un bruit qui perturbe le signal si bien que nos essais n'ont pas été concluants. D'autre part, le fait d'utiliser le Biacore pour étudier cette question, implique qu'il faut immobiliser un des anticorps ce qui introduit une gêne stérique qui n'existe pas dans le cas de l'expérience de vectorisation que l'on souhaite mener.

Récemment Bornhop *et al.* ont mis au point un système de détection (par interférométrie de réfraction de la lumière) de la liaison de deux molécules biologiques en solution ((Bornhop et al., 2007)). L'utilisation de ce type de technique permettrait d'étudier la formation d'un complexe tripartite dans des conditions s'approchant de celles de l'expérience de vectorisation.

Boulain *et al.* ont montré que la liaison d'un anticorps monoclonal augmentait la vitesse de dissociation du complexe toxine  $\alpha$  de *Naja nigricollis* / récepteur à l'acétylcholine ((Boulain and Menez, 1982), (Boulain et al., 1985)). Leurs résultats étaient en accord avec un modèle cinétique supposant l'existence d'un complexe tripartite anticorps/toxine/récepteur. La liaison de plusieurs anticorps sur la même toxine ne paraît donc pas invraisemblable.

#### **2.4.4 L'anticorps anti pTx2 permet de contrôler le repliement de Tox62**

La suite de l'étude a consisté à observer l'action de ces anticorps sur le repliement de la toxine. Etant donné la faible concentration de ces solutions d'anticorps en espèces actives, le développement d'une méthode sensible de suivi du repliement était donc nécessaire. Deux techniques basées sur des principes différents se sont avérées suffisamment sensibles pour suivre le repliement de Tox62 et Tox62 ETag à des concentrations de l'ordre du nanomolaire. La première, par mesure de l'activité biologique, est très sensible (détecte des concentrations de toxine inférieures à 1 nM) mais donne uniquement une information sur l'apparition de l'activité alors que certains intermédiaires de repliement disposent d'une activité biologique proche de celle de la molécule native. La deuxième, par isoélectrophorèse, est moins sensible mais permet de séparer et d'observer les intermédiaires de repliement. L'étude du repliement de Tox62 (ou Tox62 ETag) en présence de acpTx3 et acpTxETag par ces deux méthodes a montré que ces anticorps n'ont pas d'effet sur le repliement. A l'inverse les deux méthodes prouvent que acpTx2 a la capacité d'inhiber le repliement.

Bien que les trois anticorps reconnaissent préférentiellement la forme réduite de la toxine par rapport à sa forme oxydée, leurs effets sur le repliement diffèrent. AcpTx2 inhibe le repliement. Notons que l'interaction entre deux protéines est un phénomène dynamique. Lorsque acpTx2 et la toxine sont mis en présence, les anticorps se lient à la toxine puis peuvent ensuite se dissocier. Or, cet anticorps est capable d'inhiber le repliement pendant des temps bien supérieurs à son temps de demi-dissociation. Cela indique que le temps de recapture de l'antigène est inférieur au temps nécessaire au repliement de la partie de la molécule reconnue par l'anticorps (Figure 62 1). Les résultats du repliement en présence de acpTx2 par mesure de l'activité et par isoélectrophorèse montrent qu'au temps 24 heures environ 50% de l'activité et une quantité non négligeable de forme oxydée sont apparus. Ce résultat semble indiquer qu'une partie des toxines parviennent à s'échapper de la recapture par l'anticorps puis à se replier.

L'absence d'inhibition observée lors de l'utilisation des anticorps acpTx3 et acpTxETag peut être due à un rapport entre la vitesse de recapture et la vitesse de formation de structure dans la région de leur épitope moins favorable que dans le cas de acpTx2. (Figure 62 2). On peut également envisager que dans les régions de la toxine correspondant aux épitope pTx3 et pTxETag, la protéine est capable de se structurer et d'induire le relargage de l'anticorps (Figure 62 3). Pour obtenir des réponses plus détaillées à ces questions mécanistiques, l'obtention d'anticorps monoclonaux serait un préalable indispensable.

En plus de l'inhibition du repliement, la présence de acpTx2 provoque la formation d'un intermédiaire (probablement un intermédiaire à trois ponts) qui n'existe pas lors du repliement de Tox62 *in vitro*. AcpTx2 semble être capable de modifier la trajectoire de repliement de Tox62.

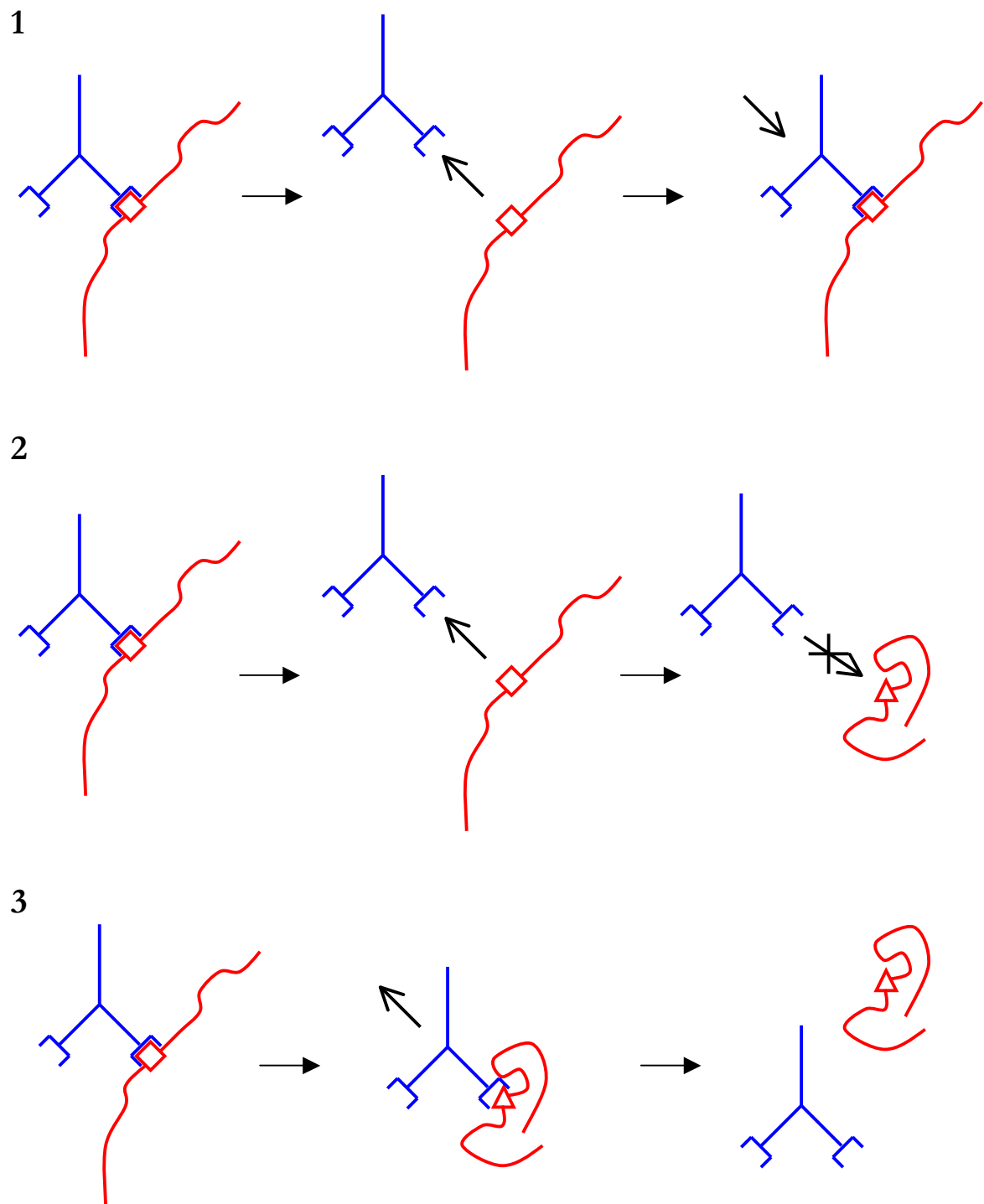


Figure 62 : Schéma interprétatif des résultats d'inhibition du repliement par les anticorps. En bleu l'anticorps, en rouge la toxine

Les résultats obtenus montrent que la stratégie adoptée pour produire des anticorps dirigés contre certaines parties de la protéine réduite est valable puisque tous les anticorps qui reconnaissent polyS reconnaissent préférentiellement la toxine réduite à la toxine oxydée. Un des anticorps (acpTx2) est capable d'inhiber le repliement et répond à l'ensemble du cahier des charges qui avait été fixé pour tenter de réaliser la vectorisation du repliement. De plus les éléments dont nous disposons indiquent qu'il est possible, même dans le cas le plus défavorable, de lier deux anticorps sur le même antigène. Il nous a manqué un deuxième anticorps (au moins, car quatre auraient été préférables) disposant des mêmes caractéristiques que acpTx2 pour pouvoir tenter une expérience de vectorisation *in vitro*.

Cependant nos résultats ne nous permettent pas de comprendre pourquoi acpTx2 inhibe le repliement alors que acpTx3 et acpTxETag en sont incapables. Pour analyser les mécanismes mis en œuvre lors de cette inhibition, il est nécessaire de disposer d'une solution homogène d'anticorps (donc d'anticorps monoclonaux) pour lesquels il est possible de définir des caractéristiques cinétiques. Dans l'état actuel de nos connaissances, il est difficile de proposer une stratégie permettant de contourner une difficulté mal identifiée. Les résultats obtenus n'excluent pas que la seule région de la molécule capable d'induire des anticorps inhibant le repliement est celle couverte par pTx2 et que toute tentative d'obtenir des anticorps inhibant le repliement en utilisant d'autres régions de la molécule est vouée à l'échec.

## 2.5 Vectorisation du repliement de protéines multidomaines

L'étude décrite ci-dessus avait pour objectif d'augmenter l'efficacité de repliement (en terme de cinétique) d'une protéine mono-domaine en « vectorisant » son repliement. Chaque anticorps masque une portion de structure secondaire et la libération séquentielle depuis l'extrémité N-terminale vers l'extrémité C-terminale doit permettre d'accélérer le repliement de ce domaine en vectorisant l'apparition des différents éléments de structure.

La même stratégie pourrait être utilisée pour tenter d'augmenter le rendement de repliement de protéines multidomaines. Le même type d'anticorps que ceux obtenus précédemment pourraient être utilisés pour vectoriser le repliement par domaine. Il serait intéressant de tenter de vectoriser le repliement d'une protéine du type de la luciférase dont le rendement de repliement *in vitro* est faible ((Frydman et al., 1999)) mais contenant des ponts disulfure (ce n'est pas le cas de la luciférase) pour que notre stratégie de vectorisation soit applicable. Si par ce biais il était possible d'augmenter le rendement de repliement de protéines multidomaines, l'importance de l'aspect vectoriel du repliement serait enfin démontrée expérimentalement.

D'autre part, la production de protéines par voies recombinantes aboutit fréquemment à des rendements de repliement faibles voir nuls ((Graslund et al., 2008)). Dans l'hypothèse où l'utilisation de ces anticorps permettrait d'augmenter le rendement de repliement d'une protéine multidomaines, il serait envisageable d'utiliser ces anticorps pour replier des protéines multidomaines qu'il est aujourd'hui impossible d'obtenir *in vitro*.

# **BIBLIOGRAPHIE**



- Adermann, K., Wattler, F., Wattler, S., Heine, G., Meyer, M., Forssmann, W. G., and Nehls, M. (1999). Structural and phylogenetic characterization of human SLURP-1, the first secreted mammalian member of the Ly-6/uPAR protein superfamily. *Protein Sci* 8, 810-819.
- Alexandrov, N. (1993). Structural argument for N-terminal initiation of protein folding. *Protein Sci* 2, 1989-1991.
- Arolas, J. L., Aviles, F. X., Chang, J. Y., and Ventura, S. (2006). Folding of small disulfide-rich proteins: clarifying the puzzle. *Trends Biochem Sci* 31, 292-301.
- Arolas, J. L., Bronsoms, S., Lorenzo, J., Aviles, F. X., Chang, J. Y., and Ventura, S. (2004). Role of kinetic intermediates in the folding of leech carboxypeptidase inhibitor. *J Biol Chem* 279, 37261-37270.
- Arolas, J. L., D'Silva, L., Popowicz, G. M., Aviles, F. X., Holak, T. A., and Ventura, S. (2005). NMR structural characterization and computational predictions of the major intermediate in oxidative folding of leech carboxypeptidase inhibitor. *Structure* 13, 1193-1202.
- Banerjee, Y., Mizuguchi, J., Iwanaga, S., and Kini, R. M. (2005). Hemextin AB complex--a snake venom anticoagulant protein complex that inhibits factor VIIa activity. *Pathophysiol Haemost Thromb* 34, 184-187.
- Basharov, M. A. (2003). Protein folding. *J Cell Mol Med* 7, 223-237.
- Bergman, L. W., and Kuehl, W. M. (1979). Formation of an Intrachain Disulfide Bond on Nascent Immunoglobulin Light-Chains. *Journal of Biological Chemistry* 254, 8869-8876.
- Bierzynski, A., Kim, P. S., and Baldwin, R. L. (1982). A Salt Bridge Stabilizes the Helix Formed by Isolated C-Peptide of Rnase-A. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* 79, 2470-2474.
- Bornhop, D. J., Latham, J. C., Kussrow, A., Markov, D. A., Jones, R. D., and Sorensen, H. S. (2007). Free-solution, label-free molecular interactions studied by back-scattering interferometry. *Science* 317, 1732-1736.
- Boulain, J. C., Fromageot, P., and Menez, A. (1985). Further evidence showing that neurotoxin-acetylcholine receptor dissociation is accelerated by monoclonal neurotoxin-specific immunoglobulin. *Mol Immunol* 22, 553-556.
- Boulain, J. C., and Menez, A. (1982). Neurotoxin-specific immunoglobulins accelerate dissociation of the neurotoxin-acetylcholine receptor complex. *Science* 217, 732-733.
- Brown, J. E., and Klee, W. A. (1971). Helix-Coil Transition of Isolated Amino Terminus of Ribonuclease. *Biochemistry* 10, 470-&.
- Bryngelson, J. D., and Wolynes, P. G. (1987). Spin-Glasses and the Statistical-Mechanics of Protein Folding. *Proceedings of the National Academy of Sciences of the United States of America* 84, 7524-7528.
- Cemazar, M., Zahariev, S., Pongor, S., and Hore, P. J. (2004). Oxidative folding of Amaranthus alpha-amylase inhibitor: disulfide bond formation and conformational folding. *J Biol Chem* 279, 16697-16705.
- Chaffotte, A. F., Guillou, Y., Goldberg, M.E. (1992). Kinetic resolution of peptide bond and side chain far-UV circular dichroism during the folding of hen egg white lysozyme. *Biochemistry* 31, 9694-9702.
- Chang, J. Y. (1996). The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry* 35, 11702-11709.
- Chang, J. Y., Canals, F., Schindler, P., Querol, E., and Aviles, F. X. (1994). The disulfide folding pathway of potato carboxypeptidase inhibitor. *J Biol Chem* 269, 22087-22094.
- Chang, J. Y., Kumar, T. K., and Yu, C. (1998). Unfolding and refolding of cardiotoxin III elucidated by reversible conversion of the native and scrambled species. *Biochemistry* 37, 6745-6751.
- Chang, J. Y., and Li, L. (2002). Pathway of oxidative folding of alpha-lactalbumin: a model for illustrating the diversity of disulfide folding pathways. *Biochemistry* 41, 8405-8413.
- Chang, J. Y., and Li, L. (2005). Divergent folding pathways of two homologous proteins, BPTI and tick anticoagulant peptide: Compartmentalization of folding intermediates and identification of kinetic traps. *Archives of Biochemistry and Biophysics* 437, 85-95.
- Chang, J. Y., Li, L., and Lai, P. H. (2001). A major kinetic trap for the oxidative folding of human epidermal growth factor. *J Biol Chem* 276, 4845-4852.
- Chang, J. Y., Lu, B. Y., Lin, C. C., and Yu, C. (2006). Fully oxidized scrambled isomers are essential and predominant folding intermediates of cardiotoxin-III. *FEBS Lett* 580, 656-660.



- Chatrenet, B., and Chang, J. Y. (1993). The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. *J Biol Chem* 268, 20988-20996.
- Cheek, S., Krishna, S. S., and Grishin, N. V. (2006). Structural classification of small, disulfide-rich protein domains. *J Mol Biol* 359, 215-237.
- Chen, W., Helenius, J., Braakman, I., and Helenius, A. (1995). Cotranslational Folding and Calnexin Binding during Glycoprotein-Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 92, 6229-6233.
- Christensen, L. L. (1997). Theoretical analysis of protein concentration determination using biosensor technology under conditions of partial mass transport limitation. *Anal Biochem* 249, 153-164.
- Christopoulos, A., and Kenakin, T. (2002). G protein-coupled receptor allostery and complexing. *Pharmacological Reviews* 54, 323-374.
- Chuang, C. C., Chen, C. Y., Yang, J. M., Lyu, P. C., and Hwang, J. K. (2003). Relationship between protein structures and disulfide-bonding patterns. *Proteins* 53, 1-5.
- Creighton, T. E. (1990). Protein folding. *Biochem J* 270, 1-16.
- Crick, F. H., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature* 192, 1227-1232.
- Daly, N. L., Clark, R. J., and Craik, D. J. (2003). Disulfide folding pathways of cystine knot proteins. Tying the knot within the circular backbone of the cyclotides. *J Biol Chem* 278, 6314-6322.
- Deane, C. M., Dong, M. Q., Huard, F. P. E., Lance, B. K., and Wood, G. R. (2007). Cotranslational protein folding - fact or fiction? *Bioinformatics* 23, I142-I148.
- Dewille, J. R., Schweitz, H., Maes, P., Tartar, A., and Lazdunski, M. (1991). Calciseptine, a Peptide Isolated from Black Mamba Venom, Is a Specific Blocker of the L-Type Calcium-Channel. *Proceedings of the National Academy of Sciences of the United States of America* 88, 2437-2440.
- Dobson, C. M., Sali, A., and Karplus, M. (1998). Protein folding: A perspective from theory and experiment. *Angewandte Chemie-International Edition* 37, 868-893.
- Drenth, J., Low, B. W., Richardson, J. S., and Wright, C. S. (1980). The toxin-agglutinin fold. A new group of small protein structures organized around a four-disulfide core. *J Biol Chem* 255, 2652-2655.
- Drevet, P., Lemaire, C., Gasparini, S., Zinn-Justin, S., Lajeunesse, E., Ducancel, F., Pinkasfeld, S., Courcon, M., Tremeau, O., Boulain, J. C., and Menez, A. (1997). High-level production and isotope labeling of snake neurotoxins, disulfide-rich proteins. *Protein Expr Purif* 10, 293-300.
- Eigenbrot, C., Randal, M., and Kossiakoff, A. A. (1990). Structural effects induced by removal of a disulfide-bridge: the X-ray structure of the C30A/C51A mutant of basic pancreatic trypsin inhibitor at 1.6 Å. *Protein Eng* 3, 591-598.
- Elcock, A. H. (2006). Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput Biol* 2, e98.
- Ellis, J. (1997). Allosteric binding sites on muscarinic receptors. *Drug Development Research* 40, 193-204.
- Epand, R. M., and Scheraga, A. (1968). Influence of Long-Range Interactions on Structure of Myoglobin. *Biochemistry* 7, 2864-&.
- Fedorov, A. N., and Baldwin, T. O. (1998). Protein folding and assembly in a cell-free expression system. *Methods Enzymol* 290, 1-17.
- Fedorov, A. N., Friguet, B., Djavadhahian, L., Alakhov, Y. B., and Goldberg, M. E. (1992). Folding on the Ribosome of Escherichia-Coli Tryptophan Synthase Beta-Subunit Nascent Chains Probed with a Conformation-Dependent Monoclonal-Antibody. *Journal of Molecular Biology* 228, 351-358.
- Fersht, A. R., and Daggett, V. (2002). Protein folding and unfolding at atomic resolution. *Cell* 108, 573-582.
- Fersht, A. R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224, 771-782.
- Friguet, B., Djavadhahian, L., King, J., and Goldberg, M. E. (1994). In-Vitro and Ribosome-Bound Folding Intermediates of P22 Tailspike Protein Detected with Monoclonal-Antibodies. *Journal of Biological Chemistry* 269, 15945-15949.
- Friguet, B., Fedorov, A., Serganov, A. A., Navon, A., and Goldberg, M. E. (1993). A Radioimmunoassay-Based Method for Measuring the True Affinity of a Monoclonal-Antibody with Trace Amounts of Radioactive Antigen - Illustration with the Products of a Cell-Free Protein-Synthesis System. *Analytical Biochemistry* 270, 344-350.

- Frydman, J., Erdjument-Bromage, H., Tempst, P., and Hartl, F. U. (1999). Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nature Structural Biology* 6, 697-705.
- Fuller, E., Green, B. R., Catlin, P., Buczek, O., Nielsen, J. S., Olivera, B. M., and Bulaj, G. (2005). Oxidative folding of conotoxins sharing an identical disulfide bridging framework. *Febs J* 272, 1727-1738.
- Galat, A., Gross, G., Drevet, P., Sato, A., and Menez, A. (2008). Conserved structural determinants in three-fingered protein domains. *Febs J* 275, 3207-3225.
- Gilbert, H. F. (1990). Molecular and cellular aspects of thiol-disulfide exchange. *Adv Enzymol Relat Areas Mol Biol* 63, 69-172.
- Gilquin, B., Bourgoin, M., Menez, R., Le Du, M. H., Servent, D., Zinn-Justin, S., and Menez, A. (2003). Motions and structural variability within toxins: implication for their use as scaffolds for protein engineering. *Protein Sci* 12, 266-277.
- Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., *et al.* (2008). Protein production and purification. *Nat Methods* 5, 135-146.
- Gupta, A., Van Vlijmen, H. W., and Singh, J. (2004). A classification of disulfide patterns and its relationship to protein structure and function. *Protein Sci* 13, 2045-2058.
- Gutte, B., and Merrifield, R. B. (1968). The total synthesis of an enzyme with ribonuclease A activity. *Journal of the American Chemical Society* 91, 501-2.
- Haber, E., and Anfinsen, C. B. (1961). Regeneration of enzyme activity by air oxidation of reduced subtilisin-modified ribonuclease. *J Biol Chem* 236, 422-424.
- Hardin, C., Eastwood, M. P., Prentiss, M., Luthey-Schulten, Z., and Wolynes, P. G. (2002). Folding funnels: The key to robust protein structure prediction. *Journal of Computational Chemistry* 23, 138-146.
- Harrison, P. M., and Sternberg, M. J. (1996). The disulphide beta-cross: from cystine geometry and clustering to classification of small disulphide-rich protein folds. *J Mol Biol* 264, 603-623.
- Hider, R.C., Drake, A.F., and Tamiya, N. (1988). An analysis of the 225-230 nm CD band of elapid toxins. *Biopolymers* 27, 113-122.
- Huard, F. P., Deane, C. M., and Wood, G. R. (2006). Modelling sequential protein folding under kinetic control. *Bioinformatics* 22, e203-210.
- Inaka, K., Taniyama, Y., Kikuchi, M., Morikawa, K., and Matsushima, M. (1991). The crystal structure of a mutant human lysozyme C77/95A with increased secretion efficiency in yeast. *J Biol Chem* 266, 12599-12603.
- Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3, 318-356.
- Karlsson, E., Mbugua, P. M., and Rodriguez-Ithurralde, D. (1984). Fasciculins, anticholinesterase toxins from the venom of the green mamba *Dendroaspis angusticeps*. *J Physiol (Paris)* 79, 232-240.
- Karlsson, R., Fagerstam, L., Nilshans, H., and Persson, B. (1993). Analysis of active antibody concentration. Separation of affinity and concentration parameters. *J Immunol Methods* 166, 75-84.
- Kauzmann, W. (1959). Some Factors in the Interpretation of Protein Denaturation. *Advances in Protein Chemistry* 14, 1-63.
- Kenyon, G. L., and Bruice, T. W. (1977). Novel sulfhydryl reagents. *Methods Enzymol* 47, 407-430.
- Kim, P. S., and Baldwin, R. L. (1990). Intermediates in the Folding Reactions of Small Proteins. *Annual Review of Biochemistry* 59, 631-660.
- Kirmizialtin, S., Ganesan, V., and Makarov, D. E. (2004). Translocation of a beta-hairpin-forming peptide through a cylindrical tunnel. *J Chem Phys* 121, 10268-10277.
- Krishna, M. M., Hoang, L., Lin, Y., and Englander, S. W. (2004). Hydrogen exchange methods to study protein folding. *Methods* 34, 51-64.
- Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science* 302, 1364-1368.
- Kumar, T. K., Jayaraman, G., Lee, C. S., Arunkumar, A. I., Sivaraman, T., Samuel, D., and Yu, C. (1997). Snake venom cardiotoxins-structure, dynamics, function and folding. *J Biomol Struct Dyn* 15, 431-463.
- Laio, A., and Micheletti, C. (2006). Are structural biases at protein termini a signature of vectorial folding? *Proteins-Structure Function and Bioinformatics* 62, 17-23.

- Laity, J. H., Lester, C. C., Shimotakahara, S., Zimmerman, D. E., Montelione, G. T., and Scheraga, H. A. (1997). Structural characterization of an analog of the major rate-determining disulfide folding intermediate of bovine pancreatic ribonuclease A. *Biochemistry* 36, 12683-12699.
- Lampugnani, M. G., Orsenigo, F., Gagliani, M. C., Tacchetti, C., and Dejana, E. (2006). Vascular endothelial cadherin controls VEGFR-2 internalization and signaling from intracellular compartments. *J Cell Biol* 174, 593-604.
- Leopold, P. E., Montal, M., and Onuchic, J. N. (1992). Protein Folding Funnels - a Kinetic Approach to the Sequence Structure Relationship. *Proceedings of the National Academy of Sciences of the United States of America* 89, 8721-8725.
- Levinthal, C. (1969). How to fold gracefully, Paper presented at: Mossbauer Spectroscopy in Biological Systems (Allerton House: University of Illinois Press ).
- Lo Conte, L., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res* 30, 264-267.
- Low, B. W., Preston, H. S., Sato, A., Rosen, L. S., Searl, J. E., Rudko, A. D., and Richardson, J. S. (1976). Three dimensional structure of erabutoxin b neurotoxic protein: inhibitor of acetylcholine receptor. *Proc Natl Acad Sci U S A* 73, 2991-2994.
- Mas, J. M., Aloy, P., Marti-Renom, M. A., Oliva, B., de Llorens, R., Aviles, F. X., and Querol, E. (2001). Classification of protein disulphide-bridge topologies. *J Comput Aided Mol Des* 15, 477-487.
- Matagne, A., and Dobson, C. M. (1998). The folding process of hen lysozyme: a perspective from the 'new view'. *Cell Mol Life Sci* 54, 363-371.
- Matouschek, A., Serrano, L., and Fersht, A. R. (1992a). The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J Mol Biol* 224, 819-835.
- Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, M., and Fersht, A. R. (1992b). The folding of an enzyme. V. H/2H exchange-nuclear magnetic resonance studies on the folding pathway of barnase: complementarity to and agreement with protein engineering studies. *J Mol Biol* 224, 837-845.
- Matthaei, H., and Nirenberg, M. W. (1961). The dependence of cell-free protein synthesis in E. coli upon RNA prepared from ribosomes. *Biochem Biophys Res Commun* 4, 404-408.
- Menez, A., Bouet, F., Guschlbauer, W., and Fromageot, P. (1980). Refolding of reduced short neurotoxins: circular dichroism analysis. *Biochemistry* 19, 4166-4172.
- Miwa, J. M., Stevens, T. R., King, S. L., Caldarone, B. J., Ibanez-Tallon, I., Xiao, C., Fitzsimonds, R. M., Pavlides, C., Lester, H. A., Picciotto, M. R., and Heintz, N. (2006). The prototoxin lynx1 acts on nicotinic acetylcholine receptors to balance neuronal activity and survival in vivo. *Neuron* 51, 587-600.
- Morrissey, M. P., Ahmed, Z., and Shakhnovich, E. I. (2004). The role of cotranslation in protein folding: a lattice model study. *Polymer* 45, 557-571.
- Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247, 536-540.
- Narayan, M., Welker, E., Wedemeyer, W. J., and Scheraga, H. A. (2000). Oxidative folding of proteins. *Acc Chem Res* 33, 805-812.
- Netzer, W. J., and Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388, 343-349.
- Nicola, A. V., Chen, W., and Helenius, A. (1999). Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nature Cell Biology* 1, 341-345.
- Nilsson, B., Moks, T., Jansson, B., Abrahmsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A., and Uhlen, M. (1987). A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng* 1, 107-113.
- Nirthanan, S., and Gwee, M. C. (2004). Three-finger alpha-neurotoxins and the nicotinic acetylcholine receptor, forty years on. *J Pharmacol Sci* 94, 1-17.
- Pardi, A., Wagner, G., and Wuthrich, K. (1983). Protein conformation and proton nuclear-magnetic-resonance chemical shifts. *Eur J Biochem* 137, 445-454.
- Pitsyn, O. B. (1987). Protein Folding - Hypotheses and Experiments. *Journal of Protein Chemistry* 6, 273-293.
- Pitsyn, O. B. (1995). Structures of Folding Intermediates. *Current Opinion in Structural Biology* 5, 74-78.
- Pitsyn, O. B., and Rashin, A. A. (1975). Model of Myoglobin Self-Organization. *Biophysical Chemistry* 3, 1-20.

- Rahman, S., Flynn, G., Aitken, A., Patel, Y., Hussain, F., Lu, X., Loftus, J. C., French, D., Wijelath, E., Strand, K., and Savidge, G. F. (2000). Differential recognition of snake venom proteins expressing specific Arg-Gly-Asp (RGD) sequence motifs by wild-type and variant integrin  $\alpha$ IIb $\beta$ 3: further evidence for distinct sites of RGD ligand recognition exhibiting negative allostery. *Biochem J* 345 Pt 3, 701-709.
- Rauk, A. (1984). Chiroptical properties of disulfides. Ab initio studies of dihydrogen disulfide and dimethyl disulfide. *Journal of the American Chemical Society* 106, 6517-6524.
- Richalet-Secordel, P. M., Rauffer-Bruyere, N., Christensen, L. L., Ofenloch-Haehnle, B., Seidel, C., and Van Regenmortel, M. H. (1997). Concentration measurement of unpurified proteins using biosensor technology under conditions of partial mass transport limitation. *Anal Biochem* 249, 165-173.
- Roder, H., and Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr Opin Struct Biol* 7, 15-28.
- Roder, H., Maki, K., Cheng, H., and Shastry, M. C. (2004). Rapid mixing methods for exploring the kinetics of protein folding. *Methods* 34, 15-27.
- Rothwarf, D. M., and Scheraga, H. A. (1993). Regeneration of bovine pancreatic ribonuclease A. 3. Dependence on the nature of the redox reagent. *Biochemistry* 32, 2690-2697.
- Roux, P., Ruoppolo M., Chaffotte, A. F., Goldberg, M. E. (1999). Comparison of the kinetics of S-S bond, secondary structure and active site formation during refolding of reduced denatured hen egg white lysozyme. *Protein science* 8, 2751-2760.
- Ruoppolo, M., Moutiez, M., Mazzeo, M. F., Pucci, P., Menez, A., Marino, G., and Quemeneur, E. (1998). The length of a single turn controls the overall folding rate of "three-fingered" snake toxins. *Biochemistry* 37, 16060-16068.
- Ruoppolo, M., Talamo, F., Pucci, P., Moutiez, M., Quemeneur, E., Menez, A., and Marino, G. (2001). Slow folding of three-fingered toxins is associated with the accumulation of native disulfide-bonded intermediates. *Biochemistry* 40, 15257-15266.
- Salamanca, S., Li, L., Vendrell, J., Aviles, F. X., and Chang, J. Y. (2003). Major kinetic traps for the oxidative folding of leech carboxypeptidase inhibitor. *Biochemistry* 42, 6754-6761.
- Schellman, J. A. (1955). The stability of hydrogen-bonded peptide structures in aqueous solution. *C R Trav Lab Carlsberg [Chim]* 29, 230-259.
- Serrano, L., Kellis, J. T., Jr., Cann, P., Matouschek, A., and Fersht, A. R. (1992a). The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* 224, 783-804.
- Serrano, L., Matouschek, A., and Fersht, A. R. (1992b). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J Mol Biol* 224, 805-818.
- Serrano, L., Matouschek, A., and Fersht, A. R. (1992c). The folding of an enzyme. VI. The folding pathway of barnase: comparison with theoretical models. *J Mol Biol* 224, 847-859.
- Servent, D. M., A. (2001). Snake neurotoxins that interact with nicotinic acetylcholine receptors, In *Handbook of neurotoxicology*, E. J. Massaro, ed. (Totowa: Humana press), pp. 385-425.
- Shimotakahara, S., Rios, C. B., Laity, J. H., Zimmerman, D. E., Scheraga, H. A., and Montelione, G. T. (1997). NMR structural analysis of an analog of an intermediate formed in the rate-determining step of one pathway in the oxidative folding of bovine pancreatic ribonuclease A: automated analysis of  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  resonance assignments for wild-type and [C65S, C72S] mutant forms. *Biochemistry* 36, 6915-6929.
- Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M., and Baldwin, R. L. (1985). Nature of the Charged-Group Effect on the Stability of the C-Peptide Helix. *Proceedings of the National Academy of Sciences of the United States of America* 82, 2349-2353.
- Stenberg, E., Persson, B., Roos, H., and Urbaniczky, C. (1991). Quantitative-Determination of Surface Concentration of Protein with Surface-Plasmon Resonance Using Radiolabeled Proteins. *Journal of Colloid and Interface Science* 143, 513-526.
- Studier, F. W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* 219, 37-44.
- Studier, F. W., and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189, 113-130.

- Svetlov, M. S., Kommer, A., Kolb, V. A., and Spirin, A. S. (2006). Effective cotranslational folding of firefly luciferase without chaperones of the Hsp70 family. *Protein Science* 15, 242-247.
- Tanford, C. (1962). Contribution of Hydrophobic Interactions to Stability of Globular Conformation of Proteins. *Journal of the American Chemical Society* 84, 4240-&.
- Tokatlidis, K., Friguet, B., Devillebonne, D., Baleux, F., Fedorov, A. N., Navon, A., Djavadiohanian, L., and Goldberg, M. E. (1995). Nascent Chains - Folding and Chaperone Interaction during Elongation on Ribosomes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 348, 89-95.
- Tremeau, O., Lemaire, C., Drevet, P., Pinkasfeld, S., Ducancel, F., Boulain, J. C., and Menez, A. (1995). Genetic-Engineering of Snake Toxins - the Functional Site of Erabutoxin-a, as Delineated by Site-Directed Mutagenesis, Includes Variant Residues. *Journal of Biological Chemistry* 270, 9362-9369.
- Tsernoglou, D., and Petsko, G. A. (1976). The crystal structure of a post-synaptic neurotoxin from sea snake at A resolution. *FEBS Lett* 68, 1-4.
- Tsetlin, V. (1999). Snake venom alpha-neurotoxins and other 'three-finger' proteins. *Eur J Biochem* 264, 281-286.
- Tsou, C. L. (1988). Folding of the nascent peptide chain into a biologically active protein. *Biochemistry* 27, 1809-1812.
- Uversky, V. N., and Fink, A. L. (2002). The chicken-egg scenario of protein folding revisited. *Febs Letters* 515, 79-83.
- van den Berg, B., Chung, E. W., Robinson, C. V., Mateo, P. L., and Dobson, C. M. (1999). The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase. *Embo J* 18, 4794-4803.
- van Mierlo, C. P., Darby, N. J., Keeler, J., Neuhaus, D., and Creighton, T. E. (1993). Partially folded conformation of the (30-51) intermediate in the disulphide folding pathway of bovine pancreatic trypsin inhibitor. 1H and 15N resonance assignments and determination of backbone dynamics from 15N relaxation measurements. *J Mol Biol* 229, 1125-1146.
- van Mierlo, C. P., Darby, N. J., Neuhaus, D., and Creighton, T. E. (1991a). (14-38, 30-51) double-disulphide intermediate in folding of bovine pancreatic trypsin inhibitor: a two-dimensional 1H nuclear magnetic resonance study. *J Mol Biol* 222, 353-371.
- van Mierlo, C. P., Darby, N. J., Neuhaus, D., and Creighton, T. E. (1991b). Two-dimensional 1H nuclear magnetic resonance study of the (5-55) single-disulphide folding intermediate of bovine pancreatic trypsin inhibitor. *J Mol Biol* 222, 373-390.
- Van Regenmortel, M. H. (2001). Antigenicity and immunogenicity of synthetic peptides. *Biologicals* 29, 209-213.
- Venhudova, G., Canals, F., Querol, E., and Aviles, F. X. (2001). Mutations in the N- and C-terminal tails of potato carboxypeptidase inhibitor influence its oxidative refolding process at the reshuffling stage. *J Biol Chem* 276, 11683-11690.
- Wang, P., and Klimov, D. K. (2008). Lattice simulations of cotranslational folding of single domain proteins. *Proteins* 70, 925-937.
- Watters, A. L., Deka, P., Corrent, C., Callender, D., Varani, G., Sosnick, T., and Baker, D. (2007). The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* 128, 613-624.
- Weissman, J. S., and Kim, P. S. (1991). Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 253, 1386-1393.
- Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* 70, 697-701.
- White, F. H., Jr. (1961). Regeneration of native secondary and tertiary structures by air oxidation of reduced ribonuclease. *J Biol Chem* 236, 1353-1360.
- Yang, Y., Wu, J., and Watson, J. T. (1999). Probing the folding pathways of long R(3) insulin-like growth factor-I (LR(3)IGF-I) and IGF-I via capture and identification of disulfide intermediates by cyanylation methodology and mass spectrometry. *J Biol Chem* 274, 37598-37604.
- Yasuda, O., Morimoto, S., Jiang, B., Kuroda, H., Kimura, T., Sakakibara, S., Fukuo, K., Chen, S., Tamatani, M., and Ogihara, T. (1994). FS2, a mamba venom toxin, is a specific blocker of the L-type calcium channels. *Artery* 21, 287-302.
- Zarrine-Afsar, A., Larson, S. M., and Davidson, A. R. (2005). The family feud: do proteins with similar structures fold via the same pathway? *Current Opinion in Structural Biology* 15, 42-49.

- Zinn-Justin, S., Roumestand, C., Gilquin, B., Bontems, F., Menez, A., and Toma, F. (1992). Three-dimensional solution structure of a curaremimetic toxin from *Naja nigricollis* venom: a proton NMR and molecular modeling study. *Biochemistry* *31*, 11335-11347.



# **ANNEXES**





---

## **ANNEXE 1**

---

Article : « The three fingered toxins (3FTs) »

Article soumis pour la publication d'un ouvrage sur l'utilisation des toxines en pharmacologie.  
Edité par Maria Elena De Lima (ICB, Universidade Federal de Minas Gerais, Brazil)



# The Three-Fingered Toxins (TFTs)

Grégori Gross<sup>1</sup>, Pascal Drevet<sup>1,‡</sup>, Denis Servent<sup>2</sup>, Renée Ménéz<sup>2</sup> & André Ménéz<sup>1,3,‡</sup>

<sup>1</sup> CEA/DSV/iBiTEC-S/SBIGeM, F-91191 Gif sur Yvette Cedex, France

<sup>2</sup> CEA,DSV/iBiTEC-S/ SIMOPRO, F-91191 Gif sur Yvette Cedex, France

<sup>3</sup> Muséum national d'Histoire naturelle, 57 Rue Cuvier, 75015 Paris, France

<sup>‡</sup> Corresponding authors

RUNNING TITLE: The Three-Fingered Toxins

## **ABSTRACT**

*Venomous glands produce a small number of protein folds that exert myriads of toxic activities. This review focuses on the three-fingered fold and the associated diversity of toxic activities. These activities include the antagonism, agonism and modulation of the function of a diversity of ligand-gated ion channels and G protein-coupled receptors. They also include inhibition of enzymes, blockage of integrins and alteration of cell membranes. The size of the three-fingered fold varies from about 60 to 80 amino acid residues with at least four disulfide bonds. The molecular modes of action of the three finger-folded toxins will be described. In particular, mutagenesis and crystallographic studies of toxin-target complexes will be presented. Also, some biochemical aspects of the three finger-folded toxins will be described, including refolding properties and evolutionary characteristics like the unusually fast rate of mutations of their genes. Finally, the review will show that the three-fingered fold can exert a diversity of other key functions unrelated to the toxic ones found in venom glands.*

## **1. INTRODUCTION**

On the 1<sup>st</sup> of October 2007, the number of structures of proteins deposited at the PDB database (1) reached the value of 42,350. Based on 34494 PDB Entries, the Structural Classification of Proteins (SCOP) database (2) indicated that the number of protein folds is 1086 (3). This relatively small value is a little surprising as an increasing number of new structures is published every year, especially since the advent of structural genomics, suggesting to Levitt that the total number of folds may level off to approximately 1600 (4), with no more than 85 of them corresponding to small proteins (1). The small number of

protein folds is now commonly considered to reflect an economy of structures that can nevertheless be associated with a prodigality of functions, as illustrated for example, in the case of animal toxins (5).

According to SCOP, the basic classification of proteins unit is a protein domain, an evolutionary unit that it is either found isolated in nature, or in different multidomain proteins (6). In 2006, it was proposed that 2945 disulfide-rich small protein domains can be classified into 41 protein folds according to structural similarity (7). According to this classification an average disulfide-rich small protein domain includes  $57 \pm 29$  residues and  $3 \pm 1$  disulfide bonds and most of these domains are from eukaryotic origin. Such domains are found in venoms where they exert a diversity of toxic functions (see Table I of reference (7)).

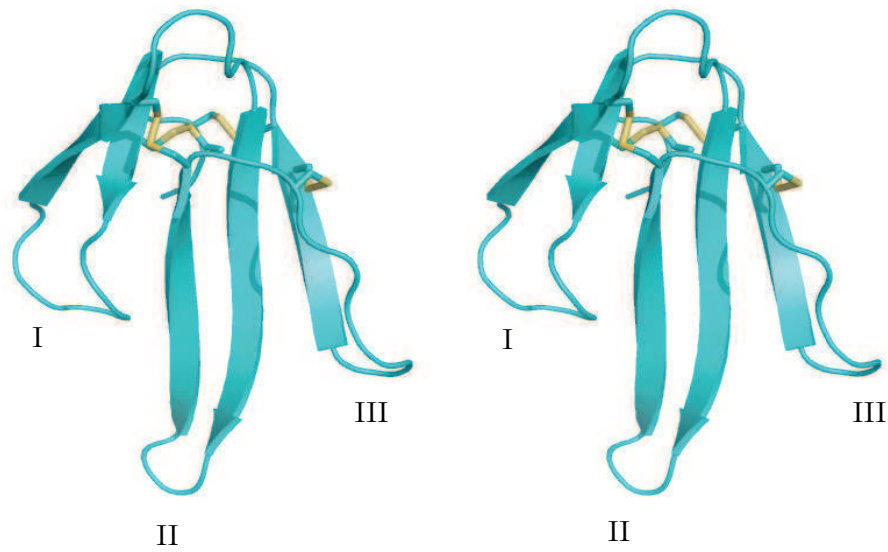
This review will focus on the disulfide-rich small protein domain that adopts a three-fingered topology. Cheek *et al* classified proteins adopting this topology as one of the families (the snake toxin-like family) forming the fold group 24 that possesses the structural core called knottin-like II. However, as it will be shown in this review, three-fingered proteins can exert many more functions than the toxic ones found in venoms, suggesting that the name “snake toxin-like family” is too limited to reflect the genuine functional potentiality of the fold. In fact, the domain can exert two major types of functions. First, the domain may act as ligands or more precisely as inhibitors, agonists or modulators of the biological activity of a physiological target (receptor, ion channel, enzyme etc.). Second, the domain may act as receptor or as building block of multimeric proteins (receptors, enzyme inhibitors, etc). This review we will primarily focus on the three-fingered proteins that exert toxic activities, and will be called Three-Fingered Toxins (TFTs).

## **2. STRUCTURAL AND EVOLUTIONARY CHARACTERISTICS OF THE THREE-FINGERED DOMAIN (TFD)**

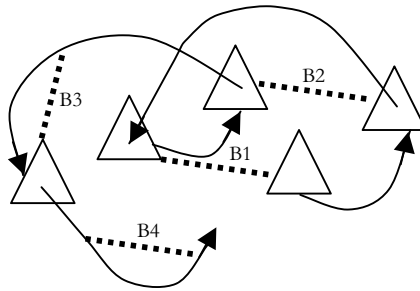
### **2.1. The Three-Fingered Domain (TFD).**

A typical Three-Fingered Domain is shown in Figure 1A. It is defined as a small disulfide-rich protein monomer whose topology comprises three loops (or fingers) emerging from a small hydrophobic core where invariant disulfides are located, with five  $\beta$ -strands forming a twisted  $\beta$ -sheet. The fold of a TFD can be represented schematically as shown in Figure 1B.

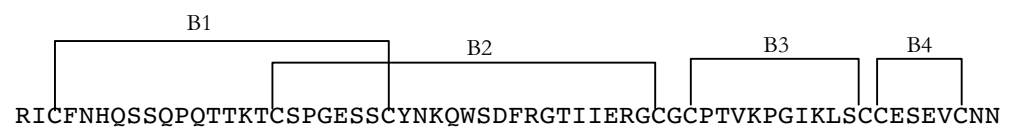
A



B



C

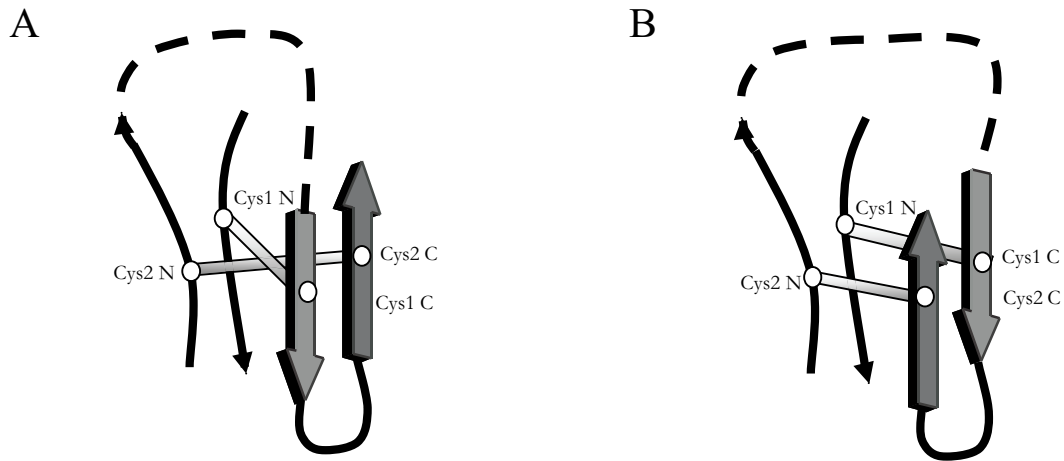


**Figure 1:** A) Stereo view of the structure of a typical Three Finger Domain. B) Organization of the secondary structure and connections of a typical TFD fold. B1-B4 correspond to the four conserved disulfide bonds. Note that B1, B2 and B4 are virtually parallel to each other, while B3 adopts a substantially different orientation. C) Amino acid sequence of erabutoxin a, a typical snake toxin adopting a TFD fold

This domain was discovered in snake venom where it exerts a potent blocking activity of the nicotinic acetylcholine receptor (8,9) . Since then, a diversity of TFTs with similar topology but different biological activities has been discovered in snake venoms. In addition, during the most recent decades other TFDs exerting key functions other than those found in venom glands were discovered within the animal kingdom. They include CD59, type II Activin Receptor, Bone Morphogenic Protein Receptor Ia and TGF- $\beta$  type II receptor. In the PFAM database TFDs were also found as building blocks of the UPAR\_Ly6 clan, which include plasmatic PLA<sub>2</sub> inhibitors and Urokinase Plasminogen Activator Receptor.

## 2.2. Structural classification of TFDs within small disulfide-rich protein domains.

Disulfides have often been considered to proceed to the structural classification of small disulfide-rich domains. Thus, classifications are usually based on either the spatial geometry of the disulfides (7,10,11) or on their sequential pattern organization (12,13). Thus, Harrison and Sternberg (10) identified a substructure called disulfide  $\beta$ -cross (Figure 2A), which contains two cystines packed against a  $\beta$  hairpin, where each C-terminal half-cystine is part of one  $\beta$  strand. This motif has an overlapping disulfide connectivity arrangement, as can be shown in figure 1C for B1 and B2 of TFDs. In addition, the authors defined the chirality of connection for the loops bridged by a disulfide. In the case of TFDs, the chain trace joining the parallel segments containing Cys2N and Cys2C follow a left-screw path via Cys1C (Figure 2B). The same authors identified another structural feature where three cystines B1, B2 and B3 are packed together. They called it the “three-cystine stack”. In the case of TFDs, the three disulfides include the two disulfides of the  $\beta$ -cross together with the C-terminal disulfide (Figure 1C). A provisional classification was thus achieved where TFDs belonged to a class described as 2 $\beta$ -hairpin/2 cystine units, with a “disulphide  $\beta$ -cross” and a three-cystine stack (Figure 1B).



**Figure 2:** A) The disulfide  $\beta$ -cross observed in small proteins rich in disulfides, as identified by Harrison and Stenberg (10). B) When the elements of  $\beta$ -cross are left connected, the spatial crossing of the bridges vanishes and two parallel bonds are observable (see text).

Mas *et al* developed an automated method, called KNOT-MATCH, to cluster proteins possessing more than three disulfides with the same spatial disulfides topology (11). The program calculates RMSD for different set of disulfides and clusters them according to a predefined threshold. This method created thirteen classes of proteins. In the TFD-containing classes, the three disulfides were approximately located on a ladder rung with a RMSD below 2.5 Å, which recalls the topology of the three-cystine stack defined by Harrison and Sternberg. This class is the most populated one, but it includes not only TFDs but also domains adopting an unrelated topology such as Nerve Growth Factor or Defensin. Thus, the above disulfide-based method of classification does not strictly reflect a TFD topology. More recently, Cheek *et al.* (7) classified the 2945 small protein entries found in PDB in 41 fold groups. Each fold group was differentiated from the others by either the architecture of secondary structure elements, topological connectivity between them, or both. In this classification scheme, TFDs were classified in a group called Knottin-like II, the expression Knottin-like corresponding in fact to another way of designing the disulfide  $\beta$ -cross. The overall Knottin-like group includes 40% of all small disulfide-rich proteins and split into three subgroups according to the connectivity between the different elements. In Knottin-like II, the four half-cystines forming the crossover are located on four structural elements whose connectivity is left-handed. This type of connection implies that the crossing over is only true in sequential term (Figure 1C). Spatially, the two disulfide bridges are in fact parallel (Figure 2B). However, the Knottin-like II group not only includes TFDs but also the topologically unrelated Hirudin-like proteins and Granuline-like repeat N-terminal domain. Therefore, the



disulfide-based criteria used by Cheek *et al.* for structural classification of small proteins rich in disulfides do not strictly reflect topological organizations of TFDs. Previously, TFDs were brought closer to the relatively similar structural family that adopts a plant lectin-like topology (like Wheat Germ Agglutinin), to form the toxin-agglutinin fold (14). In their classification, Cheek *et al.*, 2006 proposed that the two families belong to distinct group folds in agreement with a more recent study (Galat *et al.*, in preparation).

Another way to classify structurally-related disulfide-rich proteins is based on disulfides connectivity. Good correlation between protein structures and disulfide patterns was thus obtained (13), and even improved by introducing the spacing between cysteines as an additional parameter (12). Cystine number, connectivity and spacing were converted in a series of numbers named disulfide signature. Odd-numbered positions represent the spacing between pairs of cysteines forming a disulfide, and even-numbered positions indicated the number of residues separating the first-occurring cysteines of consecutive disulfides (shown in italics here after). For example, the disulfide signature of erabutoxin a (Figure 1C) is 21-14-24-26-11-12-5. Mathematical distances between two different disulfide signatures could then be easily calculated, allowing a comparison and then a clustering of disulfide-rich proteins. When five disulfides were taken into account, this method was sufficiently efficient to specifically group all the TFDs in one cluster. But in the four disulfides subdivision, the more frequent case for snake toxin TFDs, the cluster was "contaminated" by the Laminin EGF-like family. Thus, even if the half-cystines pattern appears as the most typical feature of the TFDs, they seem insufficient to define the three-fingered topology. Additional parameters remain now to be identified to obtain a better reflection of this topology.

### 2.3. Primary structure determinants of TFTs.

To understand the evolution of snake toxins adopting a TFD fold (15), a large set of amino acid sequences were aligned, revealing a strict conservation of the eight half-cystines forming the four disulfide bridges B1, B2, B3 and B4 (Figure 1C). Moreover, the disulfide bridge pattern can be used to distinguish between various nicotinic toxin families such as "short chain" neurotoxins with the four conserved disulfides, "long chain" toxins with an additional fifth disulfide at the tip of the loop 2 and "weak" neurotoxins with a fifth disulfide located in the loop 1. Furthermore, a high conservation of an asparagine vicinal to the most C-terminal half-cystine could be seen though this feature was not noticed. The only three

sequences that do not possess it correspond to molecules with unknown function and classified in two orphans groups. It should be noticed that this asparagine is included in the definition of the consensus of Activin receptor in the PFAM database suggesting the widespread character of this conserved position in TFDs, reinforcing its probable structural importance in the fold. Another constant feature is the length of the turn between the two half-cystines of the C-terminal disulfide bridge, which comprises 4 or 5 residues. This characteristic is also found in the PFAM definition of Activin receptor highlighting the wide-ranging aspect of this feature in TFPDs.

These conserved asparagine and cysteines pattern seems to be shared by all TFDs, even when they are evolutionary as distant as a sea snake neurotoxin (16) and human TGF- $\beta$  type II receptor (17). Beside this, loops and turns vary substantially in length and composition providing a high functional flexibility to this structure. Finger 1, that includes approximately a dozen residues in snake toxins (18) can be twice as long with two additional disulfide bridges in TGF- $\beta$  type II receptor (17). One of this disulfide is also present in the whole UPAR\_Ly6 family (19-21) and in some snake toxins (22-24) but not in Activin type II receptor (25,26). The presence of this topologically equivalent disulfide in evolutionary distant molecules has been considered to reflect some ancestral aspect of this element (15).

One could expect a greater conservation of Finger 2, as it corresponds to the hairpin included in the disulfide  $\beta$ -cross definition. In fact this loop is relatively constant in length with some variations in compositions, including inclusion of a disulfide bridge like in  $\alpha$ -bungarotoxin (27). Finger 3 is also highly variable in composition though relatively constant in length. There is no additional structural elements in Finger 3 among the snake toxin although an additional disulfide bridge exists in Activin type II receptor (26) and an  $\alpha$ -helix is found in CD59 (20). The last variable element is the loop that connects Finger 2 to Finger 3 that can be of one or three residues long in snake toxins, and nineteen in TGF- $\beta$  type II receptor (17), showing the large adaptability of this region without affecting the overall structure of the fold.

#### 2.4. Folding of toxins adopting a TFD fold.

In the 60's, Anfinsen and co-workers showed that when dissolved in an aqueous folding buffer, a fully denatured purified protein refolds spontaneously to its native-like

structure (28). These data demonstrated for the first time that all the information necessary for the folding of a protein is included in its sequence of amino-acids. However, there is still no definitive explanation as to the mechanism by which a denatured protein refolds to the native structure. Two models are currently prevailing (29). The first one is the framework model in which the secondary structure folds first and is followed by the formation of long range interactions between early structures. The second model is the hydrophobic collapse model in which the initial event is a disorganised compaction of the molecule, followed by a rearrangement of the structure in a confined volume. Since then, experiments proved that both these models could be valid depending on the protein studied.

The oxidative folding of small disulfide-rich proteins has been intensively studied *in vitro* during the four last decades (30). These studies show that proteins form metastable intermediates containing an incomplete number of disulfide bridges which can be purified and characterized, providing thus information on the kinetics of the reaction and on the folding pathway followed by the molecule.

Oxidative folding of several members of short-chain neurotoxins (60-62 residues and 4 disulfides) revealed that neurotoxins do not all fold at the same rate, those with 62 residues refolding more slowly than those with 60 or 61 residues (31). The difference in length, localized in the turn 2 of the molecule (see Fig.1A) was proposed to be responsible for the observed difference in kinetics, suggesting a control of turn 2 on the refolding kinetics of short-chain neurotoxins. However, as the primary structures of these proteins differ in many other positions, this conclusion needed to be supported by additional experimental work. To that goal, recombinant variants of a short toxin possessing 60, 61 or 62 residues, the additional residues being all localized in the turn 2, were prepared and their folding rates were compared in the presence of thiol catalysts (32). The data obtained confirmed the suggested correlation and thus demonstrated that the length of turn 2 indeed controls the overall folding kinetic of short-chain neurotoxins. More recently, the oxidative folding pathway of toxin  $\alpha$  62 (the variant with 62 residues) has been further explored (33), revealing the accumulation of two predominant intermediates containing three native disulfide bonds and possessing a native-like structure. Oxidation of the intermediate possessing B1, B2 and B4 (Fig. 1B) could lead to the native-like form of the toxin, while oxidation of the intermediate possessing B1, B3 and B4 did not, this intermediate appearing thus as a dead end on the folding pathway. In

the light of these results, formation of stable native-like intermediates appears to slow down the folding process and to limit its efficiency.

The oxidative folding of cardiotoxin-III, another three fingered-folded toxin was investigated *in vitro*. Its overall folding was globally comparable to that of several other non three-fingered folded proteins, including hirudin and leech carboxypeptidase inhibitor (34). Along their oxidative folding pathways these proteins accumulated a high number of heterogeneous intermediates, with a non predominance of intermediates containing only native disulfide bridges and an accumulation of fully oxidized scrambled isomers. During its oxidative folding process, cardiotoxin-III, like the two other proteins, seems to undergo a hydrophobic collapse, whereas Toxin  $\alpha$  62, which adopts also a three-fingered fold, has its oxidative process that fits better with the framework model (30). A similar conclusion is reached with the tick anticoagulant protein and the bovine pancreatic trypsin inhibitor, two proteins that share the same fold (35). Therefore, oxidative refolding processes of similarly folded proteins can follow either of the two models indicated above. It is quite intriguing that proteins sharing the same domain fold do not undergo more similar folding oxidative processes. This situation raises the question as to future researches should not be more focused on oxidative folding processes as they take place in the cell.

## 2.5. Evolution of toxins adopting a TFD fold

### 2.5.1. Amino acid sequence analyses

The origin of the toxic proteins contained in venoms remains a puzzling question. It has been observed that they display a high sequence variability, which is unexpectedly more pronounced in the coding regions of their genes as compared to their non-coding regions (36). This was particularly obvious in the regions encoding the active site of the protein. Thus it was postulated that a mechanism of accelerated evolution might occur in snake venom glands, which may be at the origin of the large sequence and functional diversity of snake toxins (36). In a detailed investigation (15), it was suggested that TFTs multigene family might evolve through a process of gene duplication and loss. The new gene created by duplication could then evolve to a new function without the loss of the first one. The duplication event could have appeared after the speciation, leaving thus a molecule characteristic of a clade, such as fasciculins in *Dendroaspis* genus. Conversely some molecules, such as "short chain" and

"long chain"  $\alpha$ -neurotoxins were found to be present in many of the examined genera, suggesting that duplication had occurred at the beginning of the elapids diversification.

According to Fry *et al*, the "birth and death" model could explain the presence of clade specific toxins by three possible scenarios. First, a gene could have been duplicated before divergence occurred and one copy has been lost in one ensuing phylum. Second, a new gene could have emerged when the duplication followed the divergence. Third, the protein could be present in other taxa without having been identified yet (15). In fact all these scenarios have been encountered in the subsequent literature. The third scenario, for example, has been validated by the characterization of a TFT in a non-venomous colubrid snake, *Coelognathus radiatus* (37). More recently an extensive study of cDNA from venom gland of *Lachesis muta* revealed the presence of TFT-like proteins in the venom of a *Viperidae* (38). These works demonstrate that TFTs are widespread, even out of the Elapidae family, and suggest an early recruitment by venomous snakes of the three-finger fold as a tool for toxic activity. The first scenario was confirmed in a study concerning the cDNA library of venom gland of *Aipysurus eydouxii* (39). This marine colubrid snake has shifted his diet from fish and eels to fish eggs. Many changes accompanied this diet modification among which a great reduction in venom toxicity. Li *et al* found that the cDNA coding for the unique TFT expressed in the venom gland shows a deletion of two nucleotides leading to an inactive form of the toxin. They postulate that the diet modification decreases the selective pressure upon the toxin which, deprived of target, has diverged. Anyhow, this loss reinforces the role of toxin for feeding rather than defense.

The construction of a phylogenic tree using the sequences of TFTs is hampered by a high degree of saturation and short sequence lengths which produce polytomies (40). Furthermore, there is a bias in the sampling of sequences, with overrepresented groups, such as long-chain and short-chain neurotoxins and type IA cytotoxins (15). Nevertheless, it was concluded that the splitting of long-chain and short-chain neurotoxins occurred early in the history of elapids (15).

As a result of a double mechanism of evolution, snake venom TFT sequences are highly heterogeneous, making it difficult to find the ancestral gene, if any. As indicated above, TFTs belong to the structural fold named "snake toxin-like" in the SCOP database. This fold also contains a variety of extracellular surface receptor such as Ly-6 (41), Urokinase

plasminogen activator receptor (19), CD59 (42), Activin receptor type II (26), Bone morphogenic protein receptor type IA (21) TGF  $\beta$  receptor type II (17), membrane associated proteins such as LYNX (43) and soluble proteins like SLURP (44). There is a low level of sequence similarity between these proteins and TFTs, except for the cysteine pattern. The eight half-cystines that belong to the bridges B1, B2, B3 and B4 (Figure 1C) are present in almost all three-fingered folded proteins.

Many of the non toxic three-fingered proteins and some of the TFTs, with lower toxicity, also possess an additional disulfide in loop 1 we called B1a (15). These 10 conserved half-cystines have been used to propose a 5 disulfide-containing protein as an ancestral gene for the TFTs (15). Recently, a gene related to that encoding LYNX, a protein that adopts a TFD fold with the additional B1a, was discovered in the brain where it exerts nicotinic acetylcholine receptor modulator activity (43). It was tempting therefore to suggest that this three-fingered protein called Lynx1 was an ancestor TFTs. However, evidence is lacking to support this possibility. On the other hand, TFD lacking B1a have also been found in ancient organism like hagfish (45) and frog (46). Thus, up to now, it is difficult to conclude if 4 disulfides is the ancestral motif of TFD or if the loss of the fifth bridge reflects some convergent evolution.

### 2.5.2. Gene organisation

Databases reveal elucidation of about fifty genes either complete or partial (including coding regions) that correspond to TFTs from elapids, but only a fraction of them has been studied in detail (47-51). All genes show the same organization with three exons split by two introns. The first exon encodes a large part of the signal sequence, the second exon codes for the end of the signal sequence and the N terminal part of the toxin up to the middle of the loop 2, and the third exon codes for the rest of the molecule. The nucleotides sequence of exon 1 and the introns are highly conserved. In contrast, the two other exons, and particularly the exon 2, are highly variable. The low divergence of introns 2 was used to trace more accurately the phylogeny of the toxins and a phylogenetic tree was built using the sequences of long-chain, short-chain and weak neurotoxins from different species (49). This tree shows an early divergence of short-chain neurotoxins as compared to the "long-chain types" (kappa, weak and long), the short-chain neurotoxins from Australian elapids forming a separate phylogenetic group. It must be noticed that the genes encoding long-chain neurotoxins

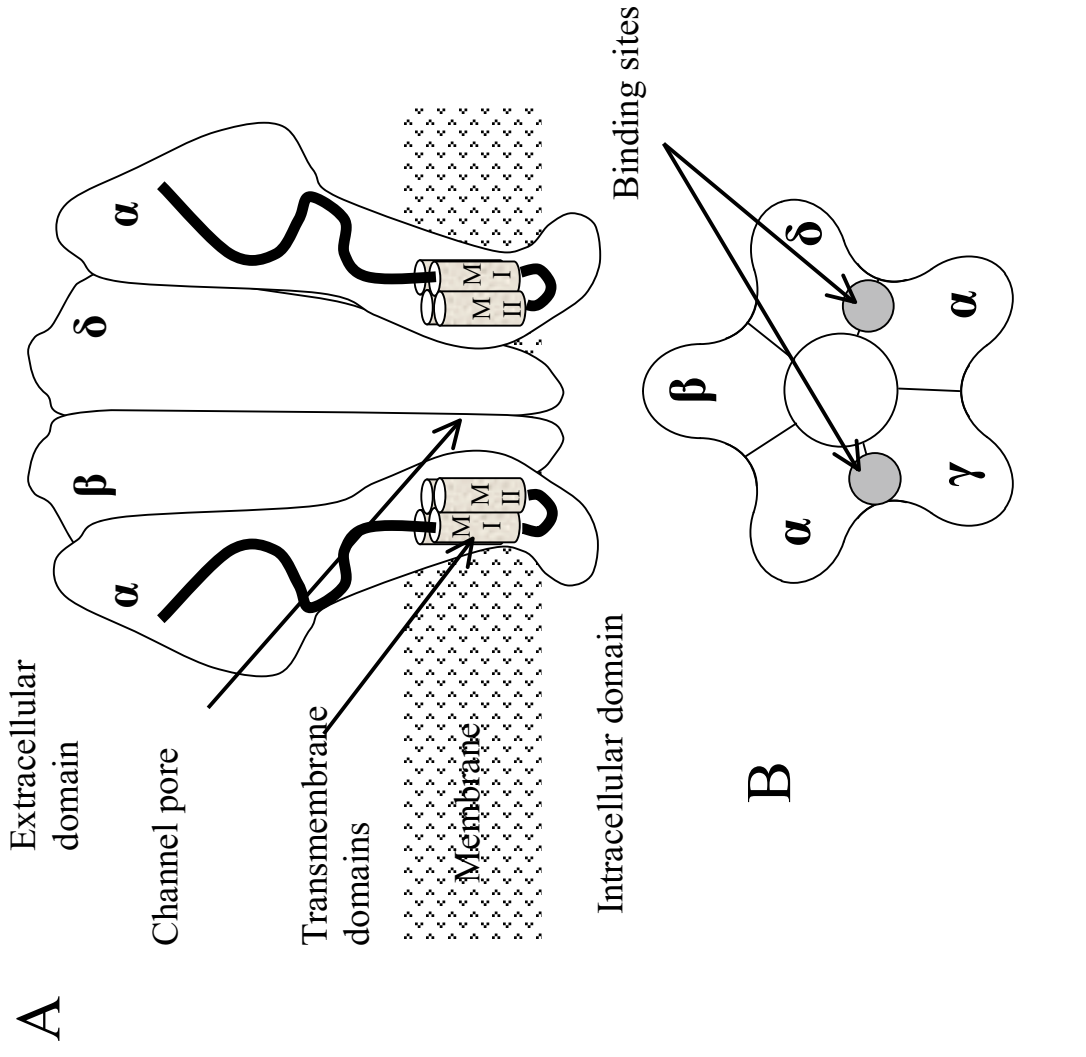
possess a specific point mutation localized in intron 2, which generates a new splicing site localized upstream of the splicing site intron 2-exon 3 observed in the genes coding for short-chain neurotoxins. This situation generates insertion of some residues at the tip of the loop 2 of "long-chain-type" neurotoxins (51). This mutation is found independently of snake species, suggesting an early event resulting in the separation between short-chain and long-chain neurotoxins (51).

Less conserved than intron 2, the sequence of intron 1 was also analyzed in detail (48). Comparing sequences of 11 genes of long and short type toxins from different snake species, it was shown that the sequence of intron 1 could be divided in ten genetic regions. Each of the sequences contains one or more of these regions, suggesting that the composition of an ancestral intron sequence has evolved by the progressive loss, and sometimes gain, of the regions. This explains the difference in length observed for the intron 1. There is no real phylogenetic tree constructed from this observation but once more the authors suggest an early separation between long-chain and short-chain toxins, each evolving later by duplication and accelerated mutation. Unlike the great variability observed for the coding sequences, introns show a relative conservation that may be more suitable to construct a robust phylogenetic tree of TFTs.

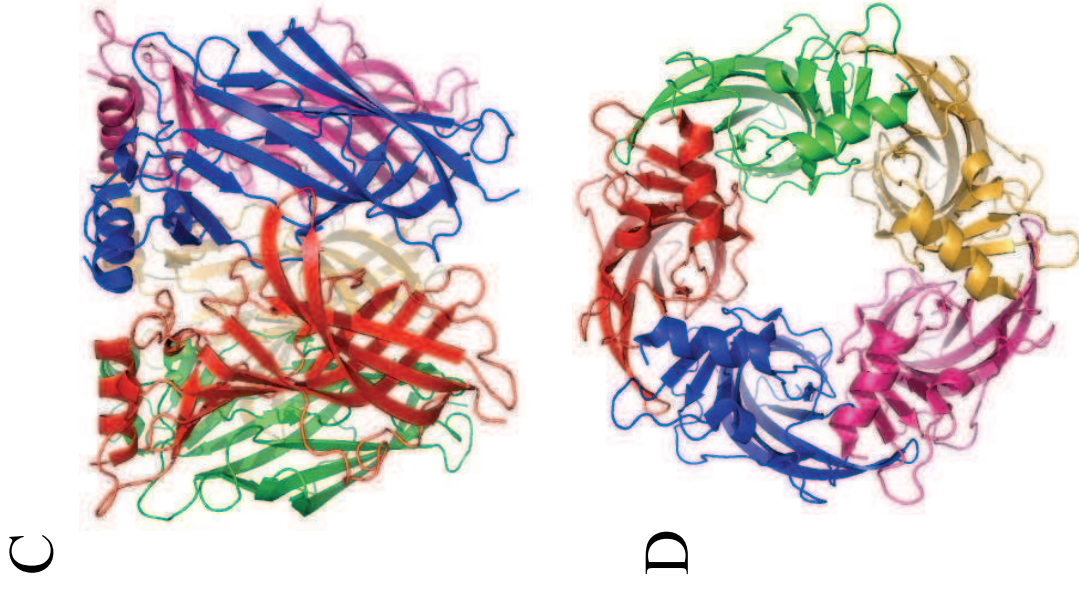
### **3. FUNCTIONS OF TOXINS ADOPTING A TFD FOLD**

#### **3.1. TFTs acting on nicotinic acetylcholine receptors (nAChRs)**

Venom peptides exhibit their toxic activity through a variety of mechanisms of actions. Among these, blockade of muscular nicotinic acetylcholine receptors by snake neurotoxins (also called curare-mimetic toxins) is probably the best characterized. These toxins bind with high affinity to nicotinic acetylcholine receptors located in the postsynaptic region of the neuromuscular (NM) junction. By maintaining the receptor in its resting state, a number of toxins cause an almost irreversible blocking of the cholinergic NM transmission (52). Toxins targeting the nAChRs show a great diversity of affinities and specificities for the different subtypes of this receptor (53,54). We will now briefly describe the classification of this receptor family and the diversity of TFTs acting on nAChRs.



**Figure 3:** Structure of the muscle nAChR. A and B: This receptor is composed of five homologous subunits in the stoichiometry  $(\alpha)_2\beta_1\gamma\delta$  in this representation. A is a longitudinal section depicting the subdomain organisation of each subunit. B is a transversal section illustrating the pentameric assembly of the molecule and the localisation of the two binding sites for agonists and competitive antagonists. C and D: X-ray structure at 2,7 Å resolution of the AChBP (PDB code : 119B) a structural homologue of the extracellular domain of nAChR, as viewed from the top (C), and perpendicular to the receptor axis (D).





### 3.1.1. Nicotinic acetylcholine receptors (nAChRs)

nAChRs are transmembrane allosteric proteins located at the neuro-muscular junction or at inter-neuronal synapses of the central and peripheral nervous system. Binding of acetylcholine to this receptor triggers a reversible opening of ionic channels, which causes a depolarisation of the postsynaptic membrane (55,56). nAChRs are pentamers, with each subunit that is constituted of one extracellular N-terminal domain that includes the binding site of agonists and antagonists, four membrane-spanning regions (MI, MII, MIII and MIV), one large and variable cytoplasmic loop and one extracellular C terminal tail (Figure 3A and B). The MII regions of each subunit form the channel pore.

nAChRs can be divided into two main families : the muscle and neuronal nAChRs. Muscle nAChRs are heteropentamers composed of  $(\alpha 1)_2\beta 1\gamma\delta$  or  $(\alpha 1)_2\beta 1\varepsilon\delta$  subunits in the embryonic or adult receptors, the embryonic-type being abundantly found in the electric organ of eels and rays (*Torpedo*). Neuronal nAChRs are either heteropentamers with various combinations of  $\alpha$  (from  $\alpha 2$  to  $\alpha 6$ ) and  $\beta$  (from  $\beta 2$  to  $\beta 4$ ) subunits or homopentamers with five identical subunits ( $\alpha 7$ ,  $\alpha 8$  or  $\alpha 9$ ) (57). Electron microscopy studies showed that nAChRs are 115 Å long. In the extracellular part, the external diameter is about 80 Å and the pore diameter 30 Å, whereas the channel is 7 Å wide (58,59). Since these pioneering structural studies, more accurate structures and structural models have been proposed. The discovery and characterisation of an acetylcholine-binding protein (AChBP) from the snail *Lymnea stagnalis* (60,61), has allowed to propose finer structural models of the pentameric assembly of extracellular domains of nAChRs. Indeed the authors crystallised the pentameric complex and obtained by X-ray diffraction a structure at a 2.7 Å resolution (Figure 3C and D). According to these data, each AChBP protomer is a single domain protein composed of an N-terminal  $\alpha$ -helix, two short  $3_{10}$  helices and a core of ten  $\beta$ -strands that form a  $\beta$  sandwich. The  $\beta$ -strands of the core are organised as in the immunoglobulin fold but here they are so twisted that the overall structure does not resemble any other protein adopting the immunoglobulin fold. The homopentamers is constituted of five of these protomers. Together they form the cylinder having the dimensions already observed by electron microscopy.

TOXIN GROUPS		NUMBER OF RESIDUES	STRUCTURAL CHARACTERISTICS	PRIMARY MOLECULAR TARGETS	OTHER TARGETS	SOURCE
<b><math>\alpha</math>-neurotoxins</b>	Short-chain	60-62	4 conserved disulfide bonds	Muscle or Torpedo ( $\alpha 1$ ) <sub>2</sub> $\beta 1\gamma\delta$ nAChR ( $K_d \approx 10^{-9} - 10^{-11}$ M)	-	<i>Elapidae</i> and <i>Hydrophidae</i> species
	Long-chain	69	4 conserved disulfide bonds	Torpedo ( $\alpha 1$ ) <sub>2</sub> $\beta 1\gamma\delta$ nAChR ( $K_d 10^{-11}$ M)	-	<i>Hydrophidae</i> ( <i>Laticauda colubrine</i> )
<b><math>\kappa</math>-neurotoxins</b>		66	4 conserved disulfide bonds + 1 extra in loop II Form dimers	Neuronal $\alpha 3\beta 2$ nAChR ( $K_d \approx 10^{-8}$ M)	Neuronal hetero and homopentamers nAChR ( $K_d \approx 10^{-8} - 10^{-6}$ M)	<i>Elapidae (kraits)</i>
<b><math>\alpha/\kappa</math>-neurotoxins</b>	Long-chain	66-74	4 conserved disulfide bonds + 1 extra in loop II	Muscle or Torpedo ( $\alpha 1$ ) <sub>2</sub> $\beta 1\gamma\delta$ nAChR ( $K_d \approx 10^{-9} - 10^{-11}$ M)	Neuronal $\alpha 7$ , $\alpha 8$ , and $\alpha 9$ homopentamers nAChR ( $K_d \approx 10^{-8}$ to $10^{-9}$ M for $\alpha 7$ )	<i>Elapidae</i> species
<b>Non-conventionnal neurotoxins</b>		65-67	4 conserved disulfide bonds + 1 extra in loop I	Muscle or Torpedo ( $\alpha 1$ ) <sub>2</sub> $\beta 1\gamma\delta$ nAChR ( $K_d > 10^{-6}$ M for WTX and $\approx 10^{-8}$ M for Candoxin)	Neuronal $\alpha 7$ nAChR for Candoxin ( $K_d \approx 5 \cdot 10^{-8}$ M)	<i>Elapidae</i> species

Table 1: Classification of three finger toxins that bind to nAChRs

### 3.1.2. The diversity of toxins acting on nAChRs

Snakes produce a diversity of toxins acting on nAChRs. They can be classified into four categories depending on the subtype of the targeted nAChR. Table 1 recapitulates this classification (53,54). The first group includes  $\alpha$ -neurotoxins (or curaremimetic toxins) that bind with high affinities to muscle or *Torpedo* nAChRs, only. It gathers many short-chain toxins of 60 to 62 residues isolated from *Elapidae* and *Hydrophidae* species and two long-chain toxins of 69 residues from *Laticauda colubrina*. They all contain the four disulfide bonds conserved in TFTs. The second group is composed of  $\kappa$ -neurotoxins that bind selectively with certain subtypes of neuronal receptors. They are long-chain toxins having an additional disulfide bond in the loop II. All of them were isolated from *kraits*. The third group comprises  $\alpha/\kappa$  neurotoxins that bind with high affinity to both muscle and neuronal homopentameric nAChRs. They are long-chain toxins with an extra bond in loop II and were isolated from *Elapidae* family. The correlation between the affinity to neuronal  $\alpha 7$  nAChRs and the presence of an extra disulfide bond in loop II suggests that this extra bridge is necessary for the binding on that subtype of receptor. This statement received an experimental support in the case of  $\alpha$ -cobratoxin from *N. kaouthia* (62). The fourth group is composed of non-conventional neurotoxins, which are short-chain toxins containing an additional disulfide in loop I. They are also called weak toxins as some studies have indicated that they can be characterised by a lower toxicity ( $LD_{50}$  in the range of 5-80 mg/kg), as compared to  $LD_{50}$  of  $\alpha$ -neurotoxins (0.04-0.3mg/kg). However, it was recently found that some of these so-called weak neurotoxins (like candoxin) could bind with high affinities to muscle and neuronal nAChRs (63,64).

### 3.1.3. Molecular aspects of toxin/nAChR interactions

#### 3.1.3.1. Binding sites of curaremimetic neurotoxins

In order to delineate the binding region of neurotoxins, amino acid sequence comparisons and chemical modifications were first achieved. Those works highlighted the functional importance of positively charged residues (lysine and arginine), aromatic residues (tryptophan and tyrosine) and the structural role of the eight conserved cysteines (65-68). The C-terminal tail of long-chain toxins was also proposed to be a functional element of

curaremimetic toxins. However, these studies were limited by the possibilities offered by the chemical modifications. A deeper exploration of the functional site of curaremimetic toxins was made possible by the emergence of recombinant techniques. The first attempt in this domain was made by isolating the cDNA encoding erabutoxin a (69) that was then expressed as a fusion protein in *E.coli* (70).

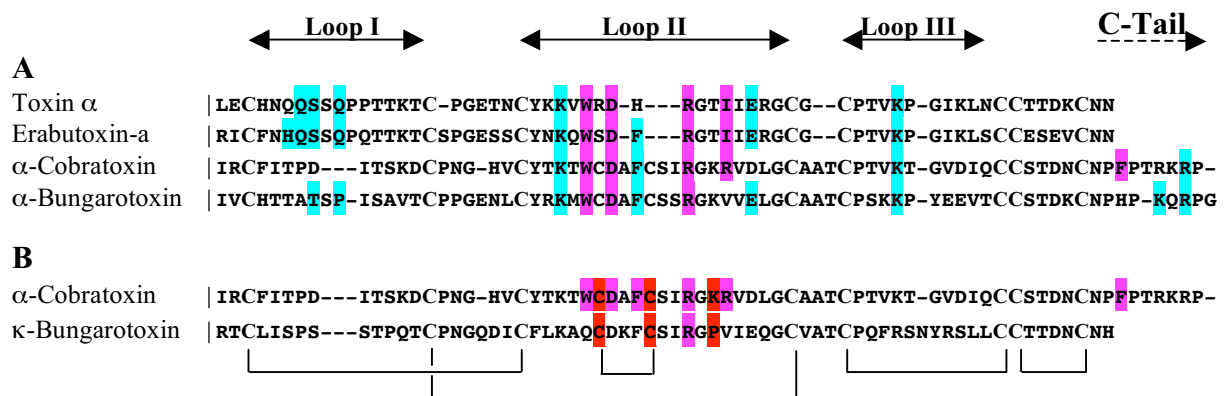
Functional site of  $\alpha$ -neurotoxins and  $\alpha/\kappa$ -neurotoxins interacting with muscle nAChRs

41 out of 62 residues of erabutoxin a, were independently mutated (71,72), and the affinity of each of the mutants for the *Torpedo* receptor was tested. The results indicated that the key amino acids by which the short-chain  $\alpha$ -neurotoxin interact with the *Torpedo* receptor (Figure 4) include residues located at the tip of loop I (Gln7, Ser8 and Gln10), on the central loop (Lys27, Trp29, Asp31 Arg33, Ile36, Glu38) and on loop III (Lys47). These results globally agreed with by other studies on NmmI from *Naja mossambica* that showed the importance of the positively charged residues at positions 27, 33 and 36 and to a lesser extent 47 (73,74). Long-chain  $\alpha/\kappa$ -neurotoxins were also submitted to site-directed mutagenesis (75-77). The most exhaustive analysis was performed on the  $\alpha$ -cobratoxin from *Naja kaouthia*. This work (75) indicated that important functional residues are localized on loop II (Lys23, Trp25, Asp27, Phe29, Arg33 and Arg36), loop III (Lys49) and on the C-terminal tail (Phe65) of the molecule. All these studies underlined the crucial role of Arg33 that was confirmed by recent studies based on the identification of receptor residues that interact with photoactivatable derivatives of a short-chain neurotoxin (78). Clearly, the determinants by which short-chain  $\alpha$ -neurotoxins and  $\alpha/\kappa$ -neurotoxins interact with muscular nAChRs share a number of similarities, suggesting that both types of proteins globally share similar binding strategies toward this subtype of receptors even if some binding specificity exist such as the critical role of the short-chain loop I in these interactions.

Functional site of  $\alpha/\kappa$ -neurotoxins and  $\kappa$ -neurotoxins interacting with neuronal  $\alpha 7$  nAChR

The same series of mutants of  $\alpha$ -cobratoxin produced for the analysis of the interaction with muscle nAChR was used to study its interaction with neuronal  $\alpha 7$  nAChR

(62) (Figure 4). Trp25, Asp27, Phe29, Arg33, Arg36, and Phe65 were shown to be important for the interaction with both types of receptor. However, Cys26-Cys30 and Lys35 were found to be specifically involved in the interaction with the neuronal receptor. Conserved residues of the dimeric  $\kappa$ -bungarotoxin were also submitted to mutations, revealing that substitutions Pro36Lys and Arg34Ala resulted in an important decrease on binding affinity to the neuronal receptor from chick ciliary ganglia (79,80). The same drastic loss of affinity was observed with the removal of the extra disulfide bond of loop II (81).



**Figure 4** : Functional residues of toxins acting on muscle (A) and neuronal (B) nAChRs. Residues specifically involved in the binding on muscle nAChRs are in blue and on neuronal nAChRs are in red. Residues implied in the binding on both receptor subtypes are in violet. Others are non tested or not implied in any interactions. See the text for references.

Taken together these data indicate that a common binding core in loop II is shared by all  $\alpha$ -neurotoxins. In particular, Arg33 at the tip of loop II appears to be a critical residue in all kinds of neurotoxins interacting with nAChRs. In addition to this conserved core, a number of residues located on loop I and loops II and III provide higher affinity and specificity toward various receptor subtypes. Furthermore, these data demonstrated the unique role played by the additional disulfide bond in loop II of long-chain  $\alpha/\kappa$ -neurotoxins and  $\kappa$ -neurotoxins in their binding with neuronal nAChRs. The surfaces by which snake toxins bind to nicotinic acetylcholine receptors therefore include two binding components, a conserved component and a specific one, a scenario that can be compared to what is also observed with other toxins that bind to voltage-gated potassium channels (82-84). Although other scenarios may also prevail, this two-components binding strategy may be spread in a variety of animal toxins.

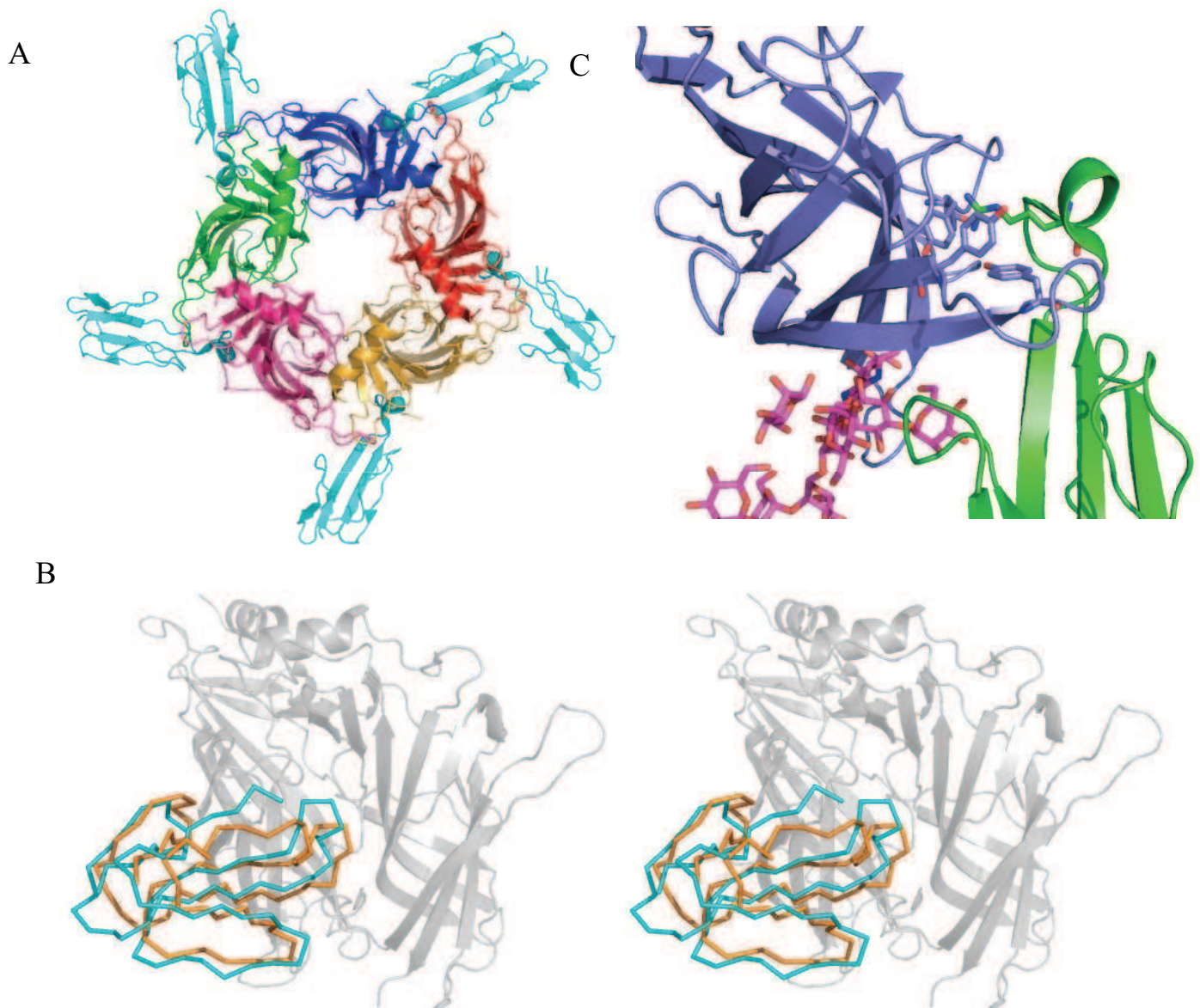
### 3.1.3.2. Binding site on nAChRs

The mapping of residues involved in the binding site of nAChRs was initially achieved using labelling probes such as MBTA (85), DDF (86,87), d-tubocurarine (88) and lophotoxin (89). The binding site was localized in a region of the  $\alpha$ -subunit composed of loop A, loop B and loop C. More recent studies showed the involvement of non- $\alpha$  subunit in the binding site of these ligands (88,90). Moreover, mutants of recombinant subunits  $\alpha$ ,  $\gamma$ ,  $\delta$  and  $\epsilon$  of murine muscle receptor were used to determine the receptor region binds to the Nmml toxin (73,91,92). Strikingly, though the binding site spread out on loops A, B, and C, only 4 mutations concentrated in loop C had an effect on the affinity. Using double-mutant analysis, the authors identified specific pairwise interactions between 3 toxin residues (Lys27, Arg33 and Lys47) and 4 residues of the  $\alpha$ -subunit (Val188, Tyr190, Pro197 and Asp200). These studies also indicated strong pairwise interactions occurred between  $\gamma$ -Glu176/Lys27,  $\gamma$ -Leu119/Arg33 and  $\gamma$ -Trp55 and Lys27 or Arg33 (91,92). Based on the same pairwise interaction approach, functional residues of  $\alpha 7$  that establish strong interactions with  $\alpha$ -cobratoxin residue side chains have been characterized. The major contacts that occurred between  $\alpha 7$  and the toxin involve the receptor loop C, which is assisted by principal (loops A and B) and complementary (loops D, F) functional regions (93).

#### 3.1.3.3. Structural insight of the interaction involving neurotoxins

##### 3.1.3.3.1. Structures of neurotoxins in complex with peptides from nAChRs

Attempts to obtain diffractable crystals of toxin-nAChRs complexes proved to be difficult for several decades. As a result, neurotoxins were first studied in complex with small peptides corresponding to various regions of the extracellular domain of nAChRs. Peptides from the loop C of the muscular  $\alpha 1$ -subunit, which included Tyr190, Tyr198 and the unique Cys192-Cys193 disulfide bond, were identified as the main part of the neurotoxin binding site (94). The same conclusion was reached with the  $\alpha 7$  nAChR. NMR and X-ray crystallography studies made with these complexes showed that the fold of the bound neurotoxin remained preserved even if slight rearrangements of the tips of loops I and II could be observed. A precise description of the molecular interaction between  $\alpha$ -bungarotoxin and the  $\alpha 1$  subunit was proposed (95).



**Figure 5** : Structure of the interaction between three finger toxins and nAChRs. A : Top view of the X-ray structure at 4,2 Å resolution (PDB code : 1YI5) of  $\alpha$ -cobratoxin in complex with nAChBP. B : Superimposition of the structure of the  $\alpha$ -cobratoxin-AChBP complex (PDB code : 1YI5, toxin in cyan) and the model of the  $\alpha$ -cobratoxin-  $\alpha$ 7 nAChR (see the text for reference, toxin in orange). The two protomers of AChBP used for the structural alignment are in grey. C : Detailed view of the X-ray structure at 1,9 Å resolution (PDB Code : 2QC1) of  $\alpha$ -bungarotoxin (in green) bound to the extracellular domain of the mouse nAChR  $\alpha$ 1 subunit (in blue and magenta for the carbohydrate chain). Finger I, finger II and the C-terminal tail of  $\alpha$ -bungarotoxin embrace the loop C of the  $\alpha$ 1 subunit. Residues implied in the cation- $\pi$  system are shown in sticks.

#### 3.1.3.3.2. Models of neurotoxins in complex with nAChRs

The crystal structure of the AChBP in complex with carbamylcholine at 2,5 Å resolution (96) confirmed the localisation of the ligand binding site, as previously deduced from photoaffinity labelling and mutagenesis experiments. The noticeable structural and functional similarities observed between the molluscan AChBP and the extracellular domain of nAChR enabled the construction of models of  $\alpha$ -neurotoxins in complex with nAChR. This was achieved by exploiting double-mutations analyses (93) or 3D structures of nAChR-peptide complexes (27,95,97). Roughly,  $\alpha$ -neurotoxins bind perpendicularly to the main axis of the cylinder defined by the five subunits of the receptor, their loop II being inserted in the ligand binding site of the receptor between two subunits (Figure 5A). The crystallographic structure of the  $\alpha$ -cobratoxin-AChBP complex not only confirmed the experimentally-based models but also highlighted the conformational changes that occurred in the receptor upon toxin binding, more particularly within the loop C (98).

#### 3.1.3.3.3. Structure of an $\alpha/\kappa$ -neurotoxin in complex with the nAChR $\alpha 1$

Recently, the crystal structure of the extracellular domain of mouse nAChR  $\alpha 1$  bound to  $\alpha$ -bungarotoxin was solved at 1.94 Å resolution (99). This structure of the  $\alpha 1$ -subunit superimposes well with the corresponding region of the  $\alpha$ -cobratoxin-AChBP complex (RMSD: 1,39 Å). The toxin is located as it was found in the structure of the  $\alpha$ -cobratoxin-AChBP complex and in the model of the cobra toxin- $\alpha 7$  neuronal nAChR complex (Figure 5B). The binding site of  $\alpha$ -bungarotoxin is precisely delineated (Figure 5C), with loop C of the  $\alpha$ -subunit being wrapped around by finger I, finger II and the C-terminal tail of the toxin. The tip of finger II inserts into the ligand binding site of the  $\alpha$ -subunit formed by loops A, B and C. Several hydrogen bonds and van der Waals contacts contribute to this interaction. At the core of the binding pocket, Arg 36 of the toxin, and Tyr 198, Tyr 190 and Tyr 93 of the  $\alpha$ -subunit, are implicated in a cation- $\pi$  interaction system, the guanidium group of Arg36 being located at the same position as the quaternary ammonium group of agonists in complex with AChBP (96). Moreover the structure shows that the finger I of  $\alpha$ -bungarotoxin interacts with the N-linked carbohydrate chain of the receptor through numerous polar contacts involving Thr6, Ala7 Thr8 and Ser9.



### 3.2. TFTs acting on G-Protein Coupled Receptors

G-Protein Coupled Receptors (GPCRs) constitute the largest superfamily of receptors, more than 850 members, accounting for 2% of the total genes encoded by the human genome. These receptors are involved in the control and regulation of all the major physiological systems, as the cardiovascular, the uro-genital, the immune, the pulmonary, the digestive and the central and peripheral nervous systems. Thus, their dysfunction is associated with lots of pathologies, explaining why this receptor family is the target of approximately 50% of the therapeutic drugs currently on the market. Nevertheless, these drugs target only 30 members of this superfamily, essentially the biogenic amine receptors. GPCRs are activated by agonists as diverse as light, ions, odorants, nucleotides, amino acids, biogenic amines, peptides and proteins. They control the activity of various effectors as enzymes and channels via their interaction with heterotrimeric G proteins (100). Despite a common seven transmembrane helices signature, these receptors have been classified in three distinct sub-families according to structural variations in the length and function of their intracellular and extracellular domains. The rhodopsin-like GPCRs family A includes the muscarinic and adrenergic receptors that are the target of the TFTs described in this review.

Five different subtypes of muscarinic acetylcholine receptor (mAChR) have been identified and classified according to their differential G-protein coupling properties. The M1, M3 and M5 mAChRs preferentially couple to  $G_q/G_{11}$  G-proteins, whereas the M2 and M4 subtypes selectively activate  $G_i/G_o$  G-proteins (101). Activation of these receptors modulates the activity of various signaling molecules, as enzymes (phospholipases, adenylate cyclase, MAP kinase) or ion channels (GIRK or voltage-activated  $Ca^{2+}$  channels), resulting finally to diverse physiological effects depending on the receptor location. The M1, M4 and M5 subtypes are mainly located in the CNS whereas the M2 and M3 are distributed both in the central and peripheral nervous system. Dysfunction of these receptors is associated with various pathologies as Alzheimer's disease, Parkinson's disease and schizophrenia for receptors distributed in the CNS or chronic obstructive pulmonary disease, overactive bladder and irritable bowel syndrome for receptors controlling the smooth-muscle contractility *via* the parasympathetic system.

Adrenergic receptors (ARs) are expressed on virtually every cell type in the body and are the receptors for epinephrine and norepinephrine, the main regulators of homeostasis. Molecular cloning has led to the identification of nine AR subtypes, which can be divided into three groups, the  $\alpha_1$ ,  $\alpha_2$ , and  $\beta$  receptors. These groups are respectively coupled to G-protein Gq, Gi and Gs and each of these groups is composed of 3 receptor subtypes:  $\alpha_{1A}$ ,  $\alpha_{1B}$ ,  $\alpha_{1D}$ ,  $\alpha_{2A}$ ,  $\alpha_{2B}$ ,  $\alpha_{2C}$ ,  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  (102). Stimulation of ARs by catecholamines within the sympathetic branch of the autonomous nervous system results in diverse physiological effects such as regulation of the vascular and smooth muscle tone, increased heart rate and bronchodilatation. In the CNS, ARs are involved for example in cognitive and learning processes. Therefore, ARs are involved in a lot of human diseases such as cardiovascular and neurodegenerative diseases or urinary tract symptoms.

On a pharmacological point of view, mAChRs and ARs are characterized by at least two distinct ligand-binding sites. Agonists and competitive antagonists bind to the orthosteric site, located inside the transmembrane domain, whereas allosteric agents induce a significant perturbation of the kinetics of binding of ligands to the primary site by interacting with an allosteric site, located more extracellularly (103,104). The effects of allosteric ligands are consistent with a ternary complex model in which the orthosteric and allosteric agents bind simultaneously to the receptor and modify each other's affinities (105,106). Considerable efforts have been made over the last decade to identify and develop novel therapeutic agents targeting the allosteric sites of GPCRs that exploit the saturable effect of allosteric agents, their potent modulation of the functional activity of endogenous ligands and their important subtype specificity (107,108).

Due to the lack of specific ligands interacting on mAChRs or ARs, in many cases it remain unclear which specific subtype of receptors are involved in mediating their various physiological functions. In fact, the only ligands that possess naturally a high selectivity in their interaction with various muscarinic or adrenergic receptors are the snake three-finger fold toxins.



### 3.2.1. Muscarinic toxins

Muscarinic neurotoxins were purified from the venom of African mambas (*Dendroaspis angusticeps* and *Dendroaspis polylepis*) and were characterized for their ability to interact with various muscarinic receptor subtypes (review in (109,110)). About ten different muscarinic toxins have been isolated and despite a high sequence homology, they exhibit clear differences in their functional activities and possess various interaction profiles for the muscarinic receptor subtypes (Figure 6). For instance, one of the most studied toxin, the MT7 toxin (also called m1-toxin1) purified from the venom of the green mamba, binds to the M<sub>1</sub> receptor subtype with a potency in the picomolar range, at least five order of magnitude higher as compared to the other receptor subtypes (111-115) (Figure 6). In addition, recent pharmacological and functional studies on the MT7-hM<sub>1</sub> interaction have indicated that MT7 (i) inhibits the ACh-stimulated [<sup>35</sup>S]GTPγS binding to M<sub>1</sub> receptor expressed in CHO cells, (ii) blocks the subsequent interaction of [<sup>3</sup>H]NMS with this receptor for several hours, and (iii) significantly decreases or increases the atropine-induced dissociation rate of [<sup>3</sup>H]NMS and [<sup>3</sup>H]ACh respectively (112,114,116,117). Therefore, MT7 acts as a highly selective antagonist of the M<sub>1</sub> receptor subtype by interacting with huge stability at its allosteric binding site. More recently, using a monoiodinated MT7 derivative and [<sup>3</sup>H]-NMS in direct and indirect binding experiments respectively, Fruchard-Gaillard *et al* determined precisely the picomolar affinity of this toxin for the hM<sub>1</sub> receptor, highlighted the strong negative cooperativity between this allosteric toxin and NMS, and showed by kinetic experiments that various allosteric agents interact allosterically with MT7 toxin (111). These results suggest that MT7, the most potent and specific allosteric modulator on M<sub>1</sub> receptor, interacts at an allosteric site that may partially overlap those previously identified for “common” and “atypical” small alkaloidic allosteric agents.

On the other hand, MT3 (m4-toxin) purified from the same venom interacts selectively on the M<sub>4</sub> receptor subtype with a nanomolar affinity, at least two orders of magnitude higher as compared to other subtypes (118-120) (Figure 6). The other toxins which have been studied possess variable affinities for the different muscarinic receptors subtypes, from nM to μM range and due to variations in the binding protocol or to the origin and purity of the toxin preparations, the pharmacological properties reported in the literature are not often homogenous (review in (109,110) (Figure 6).

On a structural point of view, NMR and X-ray structural determinations of MT2 toxin have indicated that muscarinic toxins belong to the three-finger fold structure, characteristic of several toxins acting on the cholinergic synapse (121,122). Nevertheless, our knowledge at the molecular level, on the functional site used by muscarinic toxins to interact with mAChRs is quite limited. Also poorly understood is the molecular basis that is associated with the capacity of these toxins to interact with various receptor subtypes. In addition, absolutely no structural information is available on the toxin-mAChRs complex. For the MT7-hM1 interaction, which has been the most studied complex, using baculovirus vector-Sf9 insect cell expression (123) or yeast expression in *P. pastoris* (116), recombinant MT7 was obtained but with relatively low expression level. Nevertheless, in the latter system, two mutants were constructed and expressed, allowing the identification of one MT7 residue (Phe38) involved in the stable toxin binding to M<sub>1</sub> receptor (116).

In a previous work, the high production yield of muscarinic toxins using one-step solid-phase chemical synthesis was described (112). It was showed that synthetic MT1 and MT7 toxins possess identical physicochemical and pharmacological profiles as compared to the natural toxins and the functional role of the Arg34, located at the tip of the toxin central loop, was highlighted for both toxins although they seem to recognize the hM<sub>1</sub> receptor differently (allosterically for MT7 and competitively for MT1). With the exception of the recent paper identifying two Glu and one Tyr residues in the extracellular loops of M<sub>1</sub> receptor as critical for the MT7 interaction (124), our knowledge of the interaction site used by muscarinic toxins to exploit their functional properties and specificity for various receptor subtypes remain largely unknown.

At the opposite of all the three-finger fold nicotinic toxins that interact competitively on nicotinic acetylcholine receptors, leading to a receptor blockage, muscarinic toxins possess diverse functional properties such as antagonist, allosteric modulator or agonist, even if this last property is still under debate.

Thus, even if all the muscarinic toxins are not selective for only one receptor subtype, they appear to be invaluable tools to study the pharmacology, distribution and structure-function relationships of these receptors.

### 3.2.2. Adrenergic toxins

In their preliminary work on muscarinic toxins, the Jerusalinsky and Harvey's groups studied the selectivity of the interaction of muscarinic toxins with muscarinic receptors, and performed some binding experiments on  $\alpha$ -adrenoreceptors. The first group reported that MT1 (and MT2) had no effect on the [ $^3$ H]-prazosin binding on rat cerebral cortex membranes (125), whereas Harvey *et al.* described a significant reduction (75%) of this binding by MT1 associated with a  $K_i$  of about 0.3  $\mu$ M (126). Moreover, in the same paper a similar effect was observed on membranes prepared from vas deferens, suggesting that MT1 could act postsynaptically to block  $\alpha_1$ -adrenoreceptor. Further experiments made with rabbit vas deferens preparations suggested that MT1 activates an inhibitory presynaptic muscarinic  $M_1$  receptor but also reduces the noradrenaline response to postsynaptic  $\alpha_1$ -adrenoreceptor (127).

More recently, a new toxin was isolated from the *D. angusticeps* venom that specifically acts on  $\alpha_1$ AR (N. Gilles, personal communication; Gilles N & Ménez A (2006) patent 06/01584). AdTx1 is a 65-residue peptide, reticulated by four disulfide bridges with a primary structure that is typical of a TFT. Its affinity is in the picomolar range (410 pM) for  $\alpha_1$ AR and in the micromolar range for the other subtypes. AdTx1 and prazosin do not directly compete with each other but display a reciprocal negative allosteric binding activity. AdTx1 has been conjugated with different radioactive and fluorescent probes that allowed direct binding experiments and labeling in living cells, respectively. AdTx1 represents a new therapeutic agent that could serve as lead for the development of novel drugs for treating urinary tract pathologies, like the benign prostatic hyperplasia. Recently, R. M. Kini and his colleagues isolated from the *Ophiophagus Hannah* venom a new TFT that possesses 55% sequence identity with conventional cardiotoxins from *Naja* species. However, it causes a dose-dependent decrease of heart rates *in vivo* in whole rats as well as *ex vivo* in perfused rat hearts, revealing that it induces opposite functional effects as compared to classical cardiotoxins. This cardiac effect is most probably due to a micromolar interaction of the toxin with the  $\beta_1$  and  $\beta_2$ -adrenergic receptors, justifying its name of  $\beta$ -cardiotoxin (128).

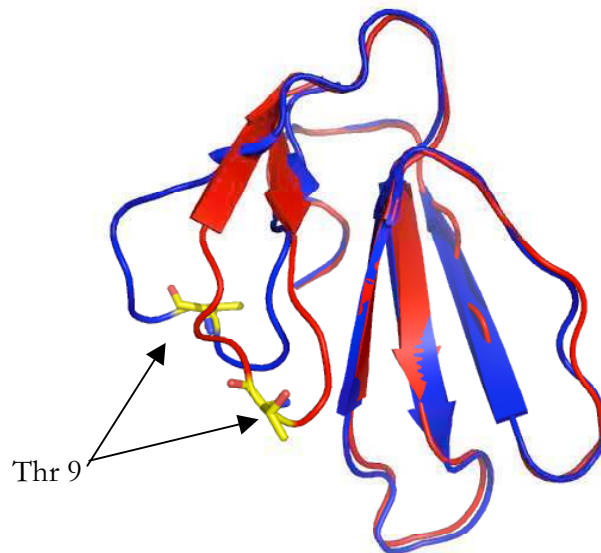
### 3.3. Other TFTs

Cytotoxicity represents another important toxic activity of elapids venom. This activity is mainly exerted by TFTs that bind to cell membranes and depolarize cardiomyocytes to cause systolic heart arrest in the envenomed victim (129). These TFTs are sixty residues long contain four disulfide bonds and show high pIs.

Two types of toxins are usually distinguished according to their lipid binding activity and have been designated as P and S type according to the presence of a proline at position 30 or a serine at position 28 respectively (130). The toxins bind preferentially to negatively charged lipids rather than neutral ones. Various proposals have been made to tentatively explain the mode of action of these toxins, including membrane disorganization, pore forming and potentialization of phospholipases (129). However, the exact mechanism of action has not yet been fully elucidated, although some interesting findings have been reported. First the interaction of the protein with lipid environment was modeled, showing that the tip of the loops could be inserted with the remaining of the toxin being maintained out by a polar "belt" (131). Furthermore, simulation data suggested differential mode of binding for the S and P type toxin, which could explain their differential binding activity (132). This model was supported by NMR experiments using  $^{31}\text{P}$  labeled lipids. Also, it was shown (133) that cardiotoxin A3 oligomerizes in the presence of charged lipids, in contrast to the non-cytotoxic cardiotoxin A5 which forms dimers only. Furthermore, cytotoxin A3 treated lipids vesicles allow the efflux of solutes but not high molecular weight proteins such as IgG. In order to observe the interaction of the toxin with charged lipids, the authors crystallized cardiotoxin A3 in the presence of SDS. The toxin/SDS molecules association confirmed the involvement of the tips of the loops in the lipid interaction. The arrangement of these toxins in the crystals have been studied, showing that the protein forms three different types of dimers that they called D1, D2 and D'2. From this intermolecular arrangement coupled with the lipid distribution they deduced a model where two tetramers of cytotoxin A3 could be inserted in a membrane, forming a pore susceptible to allow the efflux of solutes. It was proposed that the binding of the toxin on membrane lipids induces an oligomerization of the bound proteins resulting in transient pore formation responsible for the cytotoxic activity.

As mentioned above cardiotoxin A5 is devoid of any cytotoxic activity. However, it has been demonstrated that this protein possesses a strong binding activity on  $\alpha_v\beta_3$  integrin and thus may act on endothelial cells expressing this type of integral molecule (134). As cytotoxin A5 does not possess the typical RGD integrin binding motif it represents a new

mode of interaction with these integrins. In contrast, dendroaspin (or mambin), shows the typical RGD motif and binds efficiently on  $\alpha_{IIb}\beta_3$  integrin acting as a potent platelet aggregation inhibitor. This TFT from *Dendroaspis jamesonii* inhibits the binding of structure unrelated disintegrin from *Viperidae* venom. The mode of inhibition could be competitive, such as with Kistrin, or non competitive, as with Elegantin, suggesting the existence of different RGD protein binding sites on the integrin (135). However, it has been demonstrated that the RGD motif of dendroaspin could be mutated without affecting its efficiency (136). Nevertheless, the binding specificity could be modulated between  $\beta_1$  and  $\beta_3$  by replacing RGD for RYD or RCD.



**Figure 7:** Superimposition of X-ray structures of fasciculin showing the motion of loop I on binding. The bound form of the toxin is in blue and free form in red.

The *Dendroaspis* venom also contains fasciculin, another type of TFT (137). Fasciculins are very potent inhibitors of acetylcholinesterase (AChE). AChE regulates nerve impulse by rapidly catalyzing the neurotransmitter acetylcholine and its inhibition by fasciculin results in long lasting fasciculation of the victim. The molecular mechanism of action of fasciculins is relatively well known since the crystal structure of the toxin-enzyme complex has been elucidated (138,139). The catalytic site of AChE is located in a deep throat in the structure of the enzyme and the bound toxin closes the entrance of the channel



preventing the catalytic site access to the substrate. Comparison between the structure of the bound and unbound fasciculin (140) shows that the tip of loop I undergoes some movement upon binding (Figure 8). In fact, the threonine residue in position 9 is packed with a hydrophobic pocket in the soluble form of the toxin and becomes extended in the solution in the bound form. A recent modeling study on the formation of the complex suggests the possibility of an induced fit mechanism (141). Following the first contact between the enzyme and the soluble fasciculin, residues from the omega loop of AChE compete with T9 for the toxin hydrophobic pocket occupancy. Exclusion of T9 from the hydrophobic pocket results in a rearrangement of the tip of loop II leading to a new minimal energy level. It must be noticed that the rearrangement of the complex results in an opening of the "back door" of the AChE throat, which could account for the residual activity of the enzyme when bounded to fasciculin.

In addition to the well known functions described in the preceding paragraph there is a growing number of works describing new functions of TFTs. Calciseptin (142) and its homolog FS2 (143), for example, were isolated from black mamba venom and are able to specifically block L-type calcium channel resulting in a relaxation of smooth muscle and inhibition of cardiac contraction. They act as the 1,4-dihydropyridine. More recently Banerjee *et al.*, (144) isolated two TFT that they called Hemextin A and B, from an African cobra, *Hemachatus haemacatus*, which are potent anticoagulant proteins, when complexed to each other. The Hemextin AB complex is a non competitive inhibitor of the amidolytic activity of factor VIIa with a  $K_i$  of 50 nM. Interestingly, Hemextin A alone has a mild antiproteolytic activity while Hemextin B has not. Only the complex is a potent inhibitor and thus the two proteins seem to act synergistically. Where the complex is formed remains obscure since it has not been observed in the venom.

#### **4. CONCLUSION**

Toxic proteins are produced by microorganisms, plants and animals. They exert a large diversity of biological functions which surprisingly, are associated with a small number of protein folds. The present review has focused on toxic proteins that adopt one of these folds, the three-fingered fold.

We have seen that a high number of well characterized Three-Fingered Toxins (TFTs) have been identified during the past decades, most of them corresponding to monomers,

which include 60-70 residues and at least four disulfide bonds. They target a diversity of receptors, including ligand-gated receptors such as nicotinic acetylcholine receptors or L-type calcium channels and G protein-coupled receptors like muscarinic receptors or adrenergic receptors. They can also target enzymes like acetylcholinesterase or cell membrane. In fact, it has been shown that the number of biological functions associated with a Three-Fingered Fold should be substantially increased by the diversity of so-called orphan toxins, with unknown functions (15). Thus venom gland could be seen as a tank of molecules ready to use, but also ready to evolve in many different biological directions. How such an evolutionary process may occur in venom glands is not understood. During the past decade, evidence has been brought that proteins adopting other folds, like those adopting a phospholipase A<sub>2</sub> fold may undergo an accelerated rate of evolution in snake venom gland (145). Such a phenomenon has been proposed to occur also in Three-Fingered Toxin Fold (36). This may explain how the fold can adapt toward a diversity of toxic functions. However, the mechanisms associated with this evolutionary process are unclear.

Several other fundamental questions remain poorly understood. Thus, despite some *in vitro* studies, we still do not understand how Three-Fingered Toxins acquire their fold. We have seen that toxins that commonly share a three-fingered fold and the same overall biological function, i.e. blocking of the muscular nicotinic acetylcholine receptors, can refold *in vitro* at different rates. A specific  $\beta$ -turn was shown to control their differential rate of behavior and hence the production of active toxin-like form. When studying the folding of a larger diversity of three-fingered proteins it appears that one of the two major currently recognized models, i.e. the framework model and the collapse model, (29), can be adopted by some TFTs or by others, making the folding process of TFTs difficult to understand. At present, no studies have tackled the question of folding of a TFT in a genuine physiological context, i.e. within the cell venom gland. Svetlov *et al.* showed that some proteins could efficiently acquire their fold co-translationally, whereas the refolding from denatured state needs the help of chaperones (146). This suggests the importance of cotranslational folding under *in vivo* conditions. It would probably be of interest to investigate whether TFTs also refold co-translationally.

Finally, since a three-fingered fold can exert a multitude of biological functions, it is highly tempting to exploit them as a framework to develop an engineering strategy and to make them exerting new functions useful to us. A first successful attempt was achieved with the transfer of the acetylcholinesterase inhibitory activity from a fasciculin toxin to a short chain curaremimetic toxin (147,148). Furthermore recent works made on other small protein

folds rich in disulfides, have shown that this is now feasible (149,150), offering new perspectives in this domain.

Therefore, three-fingered proteins offer the possibility to tackle a multiplicity of most interesting questions in a variety of fundamental and applied directions. We suspect that they will continue to offer a remarkable domain of investigations, especially because this fold is not limited to the expression of toxic activities but to a diversity of non-toxic functions such as modulator of the nicotinic function (Lynx1, Slurp1 and 2), CD59, type II Activin Receptor, Bone Morphogenic Protein Receptor Ia and TGF- $\beta$  type II receptor, plasmatic PLA<sub>2</sub> inhibitors and Urokinase Plasminogen Activator Receptor (Galat *et al.*, in preparation).

## 5. REFERENCES

1. PDB: [http://www.rcsb.org/pdb/general\\_information/news\\_publications/newsletters/2007q3/index.html](http://www.rcsb.org/pdb/general_information/news_publications/newsletters/2007q3/index.html),
2. SCOP: <http://scop.mrc-lmb.cam.ac.uk/scop/count.html#scop-1.71>,
3. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C., 1995, J Mol Biol, 247(4), 536-540
4. Levitt, M., 2007, Proc Natl Acad Sci U S A, 104(9), 3183-3188
5. Bontems, F., Roumestand, C., Gilquin, B., Menez, A., and Toma, F., 1991, Science, 254(5037), 1521-1523
6. Lo Conte, L., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G., 2002, Nucleic Acids Res, 30(1), 264-267
7. Cheek, S., Krishna, S. S., and Grishin, N. V., 2006, J Mol Biol, 359(1), 215-237
8. Low, B. W., Preston, H. S., Sato, A., Rosen, L. S., Searl, J. E., Rudko, A. D., and Richardson, J. S., 1976, Proc Natl Acad Sci U S A, 73(9), 2991-2994
9. Tsernoglou, D., and Petsko, G. A., 1976, FEBS Lett, 68(1), 1-4
10. Harrison, P. M., and Sternberg, M. J., 1996, J Mol Biol, 264(3), 603-623
11. Mas, J. M., Aloy, P., Marti-Renom, M. A., Oliva, B., de Llorens, R., Aviles, F. X., and Querol, E., 2001, J Comput Aided Mol Des, 15(5), 477-487
12. Gupta, A., Van Vlijmen, H. W., and Singh, J., 2004, Protein Sci, 13(8), 2045-2058
13. Chuang, C. C., Chen, C. Y., Yang, J. M., Lyu, P. C., and Hwang, J. K., 2003, Proteins, 53(1), 1-5
14. Drenth, J., Low, B. W., Richardson, J. S., and Wright, C. S., 1980, J Biol Chem, 255(7), 2652-2655
15. Fry, B. G., Wuster, W., Kini, R. M., Brusica, V., Khan, A., Venkataraman, D., and Rooney, A. P., 2003, J Mol Evol, 57(1), 110-129
16. Saludjian, P., Prange, T., Navaza, J., Menez, R., Guilloteau, J. P., Ries-Kautt, M., and Ducruix, A., 1992, Acta Crystallogr B, 48 (Pt 4), 520-531
17. Boesen, C. C., Radaev, S., Motyka, S. A., Patamawenu, A., and Sun, P. D., 2002, Structure, 10(7), 913-919
18. Gilquin, B., Bourgoïn, M., Menez, R., Le Du, M. H., Servent, D., Zinn-Justin, S., and Menez, A., 2003, Protein Sci, 12(2), 266-277
19. Llinas, P., Le Du, M. H., Gardsvoll, H., Dano, K., Ploug, M., Gilquin, B., Stura, E. A., and Menez, A., 2005, Embo J, 24(9), 1655-1663

20. Kieffer, B., Driscoll, P. C., Campbell, I. D., Willis, A. C., van der Merwe, P. A., and Davis, S. J., 1994, *Biochemistry*, 33(15), 4471-4482
21. Keller, S., Nickel, J., Zhang, J. L., Sebald, W., and Mueller, T. D., 2004, *Nat Struct Mol Biol*, 11(5), 481-488
22. Pawlak, J., Mackessy, S. P., Fry, B. G., Bhatia, M., Mourier, G., Fruchart-Gaillard, C., Servent, D., Menez, R., Stura, E., Menez, A., and Kini, R. M., 2006, *J Biol Chem*, 281(39), 29030-29041
23. Kuhn, P., Deacon, A. M., Comoso, S., Rajaseger, G., Kini, R. M., Uson, I., and Kolatkar, P. R., 2000, *Acta Crystallogr D Biol Crystallogr*, 56(Pt 11), 1401-1407
24. Paaventhnan, P., Joseph, J. S., Nirthan, S., Rajaseger, G., Gopalakrishnakone, P., Kini, R. M., and Kolatkar, P. R., 2003, *Acta Crystallogr D Biol Crystallogr*, 59(Pt 3), 584-586
25. Greenwald, J., Fischer, W. H., Vale, W. W., and Choe, S., 1999, *Nat Struct Biol*, 6(1), 18-22
26. Greenwald, J., Vega, M. E., Allendorph, G. P., Fischer, W. H., Vale, W., and Choe, S., 2004, *Mol Cell*, 15(3), 485-489
27. Harel, M., Kasher, R., Nicolas, A., Guss, J. M., Balass, M., Fridkin, M., Smit, A. B., Brejc, K., Sixma, T. K., Katchalski-Katzir, E., Sussman, J. L., and Fuchs, S., 2001, *Neuron*, 32(2), 265-275
28. Anfinsen, C. B., 1973, *Science*, 181(96), 223-230
29. Daggett, V., and Fersht, A. R., 2003, *Trends Biochem Sci*, 28(1), 18-25
30. Arolas, J. L., Aviles, F. X., Chang, J. Y., and Ventura, S., 2006, *Trends Biochem Sci*, 31(5), 292-301
31. Menez, A., Bouet, F., Guschlbauer, W., and Fromageot, P., 1980, *Biochemistry*, 19(18), 4166-4172
32. Ruoppolo, M., Moutiez, M., Mazzeo, M. F., Pucci, P., Menez, A., Marino, G., and Quemeneur, E., 1998, *Biochemistry*, 37(46), 16060-16068
33. Ruoppolo, M., Talamo, F., Pucci, P., Moutiez, M., Quemeneur, E., Menez, A., and Marino, G., 2001, *Biochemistry*, 40(50), 15257-15266
34. Chang, J. Y., Kumar, T. K., and Yu, C., 1998, *Biochemistry*, 37(19), 6745-6751
35. Chang, J. Y., 1996, *Biochemistry*, 35(36), 11702-11709
36. Ohno, M., Menez, R., Ogawa, T., Danse, J. M., Shimohigashi, Y., Fromen, C., Ducancel, F., Zinn-Justin, S., Le Du, M. H., Boulain, J. C., Tamiya, T., and Menez, A., 1998, *Prog Nucleic Acid Res Mol Biol*, 59, 307-364
37. Fry, B. G., Lumsden, N. G., Wuster, W., Wickramaratna, J. C., Hodgson, W. C., and Kini, R. M., 2003, *J Mol Evol*, 57(4), 446-452
38. Junqueira-de-Azevedo, I. L., Ching, A. T., Carvalho, E., Faria, F., Nishiyama, M. Y., Jr., Ho, P. L., and Diniz, M. R., 2006, *Genetics*, 173(2), 877-889
39. Li, M., Fry, B. G., and Kini, R. M., 2005, *J Mol Evol*, 60(1), 81-89
40. Fry, B. G., 2005, *Genome Res*, 15(3), 403-420
41. Rosel, M., Claas, C., Seiter, S., Herlevsen, M., and Zoller, M., 1998, *Oncogene*, 17(15), 1989-2002
42. Fletcher, C. M., Harrison, R. A., Lachmann, P. J., and Neuhaus, D., 1994, *Structure*, 2(3), 185-199
43. Miwa, J. M., Stevens, T. R., King, S. L., Caldarone, B. J., Ibanez-Tallon, I., Xiao, C., Fitzsimonds, R. M., Pavlides, C., Lester, H. A., Picciotto, M. R., and Heintz, N., 2006, *Neuron*, 51(5), 587-600
44. Adermann, K., Wattler, F., Wattler, S., Heine, G., Meyer, M., Forssmann, W. G., and Nehls, M., 1999, *Protein Sci*, 8(4), 810-819

45. dos Remedios, N. J., Ramsland, P. A., Hook, J. W., and Raison, R. L., 1999, *Dev Comp Immunol*, 23(1), 1-14
46. Kolbe, H. V., Huber, A., Cordier, P., Rasmussen, U. B., Bouchon, B., Jaquinod, M., Vlasak, R., Delot, E. C., and Kreil, G., 1993, *J Biol Chem*, 268(22), 16458-16464
47. Jeyaseelan, K., Poh, S. L., Nair, R., and Armugam, A., 2003, *FEBS Lett*, 553(3), 333-341
48. Tamiya, T., and Fujimi, T. J., 2006, *Mol Divers*, 10(4), 529-543
49. Phui Yee, J. S., Nanling, G., Afifiyan, F., Donghui, M., Siew Lay, P., Armugam, A., and Jeyaseelan, K., 2004, *Biochimie*, 86(2), 137-149
50. Afifiyan, F., Armugam, A., Tan, C. H., Gopalakrishnakone, P., and Jeyaseelan, K., 1999, *Genome Res*, 9(3), 259-266
51. Fujimi, T. J., Nakajyo, T., Nishimura, E., Ogura, E., Tsuchiya, T., and Tamiya, T., 2003, *Gene*, 313, 111-118
52. Lee, C. Y., 1972, *Annu Rev Pharmacol*, 12, 265-286
53. Servent, D., 2001, Snake neurotoxins that interact with nicotinic acetylcholine receptors. In: Massaro, E. J. (ed). *Handbook of neurotoxinology*, Humana press, Totowa
54. Nirthanan, S., and Gwee, M. C., 2004, *J Pharmacol Sci*, 94(1), 1-17
55. Corringer, P. J., Le Novère, N., and Changeux, J. P., 2000, *Annu. Rev. Pharmacol. Toxicol.*, 40, 431-458
56. Grutter, T., and Changeux, J. P., 2001, *Trends Biochem. Sci.*, 26(8), 459-463.
57. Lindstrom, J., 1999, *Purification and cloning of nicotinic acetylcholine receptors*, Wiley-Liss, New York
58. Unwin, N., 1993, *J Mol. Biol*, 229(4), 1101-1124
59. Miyazawa, A., Fujiyoshi, Y., Stowell, M., and Unwin, N., 1999, *J. Mol. Biol.*, 288(4), 765-786
60. Smit, A. B., Syed, N. I., Schaap, D., van Minnen, J., Klumperman, J., Kits, K. S., Lodder, H., van der Schors, R. C., van Elk, R., Sorgedrager, B., Brejc, K., Sixma, T. K., and Geraerts, W. P., 2001, *Nature*, 411, 261-268
61. Brejc, K., van Dijk, W. J., Klaassen, R. M., Schuurmans, M., van der Oost, J., Smit, A. B., and Sixma, T. K., 2001, *Nature*, 411, 269-276
62. Antil-Delbeke, S., Gaillard, C., Tamiya, T., Corringer, P. J., Changeux, J. P., Servent, D., and Menez, A., 2000, *J Biol Chem*, 275(38), 29594-29601
63. Nirthanan, S., Charpantier, E., Gopalakrishnakone, P., Gwee, M. C., Khoo, H. E., Cheah, L. S., Bertrand, D., and Kini, R. M., 2002, *J Biol Chem*, 277(20), 17811-17820
64. Nirthanan, S., Charpantier, E., Gopalakrishnakone, P., Gwee, M. C., Khoo, H. E., Cheah, L. S., Kini, R. M., and Bertrand, D., 2003, *Br J Pharmacol*, 139(4), 832-844
65. Martin, B. M., Chibber, B. A., and Maelicke, A., 1983, *J Biol Chem*, 258(14), 8714-8722
66. Lin, S. R., and Chang, C. C., 1991, *Toxicon*, 29(8), 937-950
67. Faure, G., Boulain, J. C., Bouet, F., Montenay-Garestier, T., Fromageot, P., and Menez, A., 1983, *Biochemistry*, 22(9), 2068-2076
68. Chicheportiche, R., Vincent, J. P., Kopeyan, C., Schweitz, H., and Lazdunski, M., 1975, *Biochemistry*, 14(10), 2081-2091
69. Tamiya, T., Lamouroux, A., Julien, J. F., Grima, B., Mallet, J., Fromageot, P., and Menez, A., 1985, *Biochimie*, 67(2), 185-189
70. Ducancel, F., Boulain, J. C., Tremeau, O., and Menez, A., 1989, *Protein Eng*, 3(2), 139-143
71. Tremeau, O., Lemaire, C., Drevet, P., Pinkasfeld, S., Ducancel, F., Boulain, J. C., and Menez, A., 1995, *J Biol Chem*, 270(16), 9362-9369

72. Pillet, L., Tremeau, O., Ducancel, F., Drevet, P., Zinn-Justin, S., Pinkasfeld, S., Boulain, J. C., and Menez, A., 1993, *J Biol Chem*, 268(2), 909-916
73. Ackermann, E. J., Ang, E. T., Kanter, J. R., Tsigelny, I., and Taylor, P., 1998, *J Biol Chem*, 273(18), 10958-10964
74. Taylor, P., Osaka, H., Molles, B. E., Sugiyama, N., Marchot, P., Ackermann, E. J., Malany, S., McArdle, J. J., Sine, S. M., and Tsigelny, I., 1998, *J Physiol Paris*, 92(2), 79-83
75. Antil, S., Servent, D., and Menez, A., 1999, *J Biol Chem*, 274(49), 34851-34858
76. Rosenthal, J. A., Hsu, S. H., Schneider, D., Gentile, L. N., Messier, N. J., Vaslet, C. A., and Hawrot, E., 1994, *J Biol Chem*, 269(15), 11178-11185
77. Rosenthal, J. A., Levandoski, M. M., Chang, B., Potts, J. F., Shi, Q. L., and Hawrot, E., 1999, *Biochemistry*, 38(24), 7847-7855
78. Kessler, P., Thai, R., Beau, F., Tarride, J. L., and Menez, A., 2006, *Bioconjug Chem*, 17(6), 1482-1491
79. Fiordalisi, J. J., Al-Rabiee, R., Chiapinelli, V. A., and Grant, G. A., 1994, *Biochemistry*, 33, 3872-3877
80. Fiordalisi, J. J., al-Rabiee, R., Chiappinelli, V. A., and Grant, G. A., 1994, *Biochemistry*, 33(44), 12962-12967
81. Grant, G. A., Luetje, C. W., Summers, R., and Xu, X. L., 1998, *Biochemistry*, 37(35), 12166-12171
82. Dauplais, M., Lecoq, A., Song, J., Cotton, J., Jamin, N., Gilquin, B., Roumestand, C., Vita, C., de Medeiros, C. L., Rowan, E. G., Harvey, A. L., and Menez, A., 1997, *J Biol Chem*, 272(7), 4302-4309
83. Jacobsen, R. B., Koch, E. D., Lange-Malecki, B., Stocker, M., Verhey, J., Van Wagoner, R. M., Vyazovkina, A., Olivera, B. M., and Terlau, H., 2000, *J Biol Chem*, 275(32), 24639-24644
84. Gilquin, B., Racape, J., Wrisch, A., Visan, V., Lecoq, A., Grissmer, S., Menez, A., and Gasparini, S., 2002, *J Biol Chem*, 277(40), 37406-37413
85. Kao, P. N., and Karlin, A., 1986, *J Biol Chem*, 261(18), 8085-8088
86. Dennis, M., Giraudat, J., Kotzyba-Hibert, F., Goeldner, M., Hirth, C., Chang, J. Y., Lazure, C., Chretien, M., and Changeux, J. P., 1988, *Biochemistry*, 27(7), 2346-2357
87. Galzi, J. L., Revah, F., Black, D., Goeldner, M., Hirth, C., and Changeux, J. P., 1990, *J Biol Chem*, 265(18), 10430-10437
88. Chiara, D. C., and Cohen, J. B., 1997, *J Biol Chem*, 272(52), 32940-32950
89. Abramson, S. N., Culver, P., Kline, T., Li, Y., Guest, P., Gutman, L., and Taylor, P., 1988, *J Biol Chem*, 263(34), 18568-18573
90. Chiara, D. C., Middleton, R. E., and Cohen, J. B., 1998, *FEBS Lett*, 423(2), 223-226
91. Osaka, H., Malany, S., Kanter, J. R., Sine, S. M., and Taylor, P., 1999, *J Biol Chem*, 274(14), 9581-9586
92. Osaka, H., Malany, S., Molles, B. E., Sine, S. M., and Taylor, P., 2000, *J Biol Chem*, 275(8), 5478-5484
93. Fruchart-Gaillard, C., Gilquin, B., Antil-Delbeke, S., Le Novere, N., Tamiya, T., Corringier, P. J., Changeux, J. P., Menez, A., and Servent, D., 2002, *Proc Natl Acad Sci U S A*, 99(5), 3216-3221
94. Katchalski-Katzir, E., Kasher, R., Balass, M., Scherf, T., Harel, M., Fridkin, M., Sussman, J. L., and Fuchs, S., 2003, *Biophys Chem*, 100(1-3), 293-305
95. Zeng, H., Moise, L., Grant, M. A., and Hawrot, E., 2001, *J Biol Chem*, 276(25), 22930-22940
96. Celie, P. H., van Rossum-Fikkert, S. E., van Dijk, W. J., Brejc, K., Smit, A. B., and Sixma, T. K., 2004, *Neuron*, 41(6), 907-914

97. Moise, L., Piserchio, A., Basus, V. J., and Hawrot, E., 2002, *J Biol Chem*, 277(14), 12406-12417
98. Bourne, Y., Talley, T. T., Hansen, S. B., Taylor, P., and Marchot, P., 2005, *Embo J*, 24(8), 1512-1522
99. Dellisanti, C. D., Yao, Y., Stroud, J. C., Wang, Z. Z., and Chen, L., 2007, *Nat Neurosci*, 10(8), 953-962
100. Bockaert, J., and Pin, J. P., 1999, *EMBO J.*, 18, 1723-1729
101. Caulfield, M. P., and Birdsall, N. J., 1998, *Pharmacol Rev*, 50(2), 279-290
102. Bylund, D., Bond, R., Bouvier, M., Clarke, D. E., Eikenburg, D. C., Hieble, J. P., Kobika, B. K., Langer, S. Z., Lefkowitz, R. J., Minneman, K. P., and Ruffolo, R. R., 2000, *Adrenoceptors*, IUPHAR Media, London
103. Ellis, J., 1997, *Drug Dev. Res.*, 40, 193-204
104. Christopoulos, A., and Kenakin, T., 2002, *Pharmacol. Rev.*, 54(2), 323-374
105. Ehlert, F. J., 1988, *Mol. Pharmacol.*, 33(2), 187-194
106. Lazareno, S., and Birdsall, N. J., 1995, *Mol. Pharmacol.*, 48(2), 362-378
107. May, L. T., and Christopoulos, A., 2003, *Curr Opin Pharmacol*, 3(5), 551-556
108. May, L. T., Avlani, V. A., Sexton, P. M., and Christopoulos, A., 2004, *Curr Pharm Des*, 10(17), 2003-2013
109. Bradley, K. N., 2000, *Pharmacol. Ther.*, 85(2), 87-109
110. Karlsson, E., Jolkonen, M., Mulugeta, E., Onali, P., and Adem, A., 2000, *Biochimie*, 82, 793-806
111. Fruchart-Gaillard, C., Mourier, G., Marquer, C., Menez, A., and Servent, D., 2006, *Mol Pharmacol*, 69(5), 1641-1651
112. Mourier, G., Dutertre, S., Fruchart-Gaillard, C., Ménez, A., and Servent, D., 2003, *Mol. Pharmacol.*, 63, 26-35
113. Max, S. I., Liang, J. S., and Potter, L. T., 1993, *J Neurosci*, 13(10), 4293-4300
114. Olanas, M. C., Maullu, C., Adem, A., Mulugeta, E., Karlsson, E., and Onali, P., 2000, *Br. J. Pharmacol.*, 131(3), 447-452
115. Carsi, J. M., and Potter, L. T., 2000, *Toxicon*, 38(2), 187-198
116. Krajewski, J. L., Dickerson, I. M., and Potter, L. T., 2001, *Mol. Pharmacol.*, 60(4), 725-731
117. Olanas, M. C., Adem, A., Karlsson, E., and Onali, P., 2004, *Eur. J. Pharmacol.*, 487(1-3), 65-72
118. Jolkonen, M., van Giersbergen, P. L., Hellman, U., Wernstedt, C., and Karlsson, E., 1994, *FEBS Lett*, 352(1), 91-94
119. Jerusalinsky, D., Kornisiuk, E., Alfaro, P., Quillfeldt, J., Alonso, M., Verde, E. R., Cervenansky, C., and Harvey, A., 1998, *Neuroreport*, 9(7), 1407-1411
120. Liang, J. S., Carsi-Gabrenas, J., Krajewski, J. L., McCafferty, J. M., Purkerson, S. L., Santiago, M. P., Strauss, W. L., Valentine, H. H., and Potter, L. T., 1996, *Toxicon*, 34(11-12), 1257-1267
121. Ségalas, I., Roumestand, C., Zinn-Justin, S., Gilquin, B., Ménez, R., Ménez, A., and Toma, F., 1995, *Biochemistry*, 34(4), 1248-1260
122. Ménez, R., and Ducruix, A., 1993, *J. Mol. Biol.*, 232(3), 997-998
123. Nasman, J., Jolkonen, M., Ammoun, S., Karlsson, E., and Akerman, K. E., 2000, *Biochem. Biophys. Res. Commun.*, 271(2), 435-439
124. Kukkonen, A., Perakyla, M., Akerman, K. E., and Nasman, J., 2004, *J Biol Chem*, 279, 50923-50929
125. Jerusalinsky, D., Cervenasky, C., Pena, C., Raskovsky, S., and Dajas, F., 1992, *Neurochem Int*, 20(2), 237-246

126. Harvey, A., Cervenansky, C., Kornisiuk, E., and Jerusalinsky, D., 1995, *Life Science*, 56(11-12), 1014
127. Harvey, A. L., Kornisiuk, E., Bradley, K. N., Cervenansky, C., Duran, R., Adrover, M., Sanchez, G., and Jerusalinsky, D., 2002, *Neurochem Res*, 27(11), 1543-1554
128. Rajagopalan, N., Pung, Y. F., Zhu, Y. Z., Wong, P. T., Kumar, P. P., and Kini, R. M., 2007, *Faseb J*,
129. Kumar, T. K., Jayaraman, G., Lee, C. S., Arunkumar, A. I., Sivaraman, T., Samuel, D., and Yu, C., 1997, *J Biomol Struct Dyn*, 15(3), 431-463
130. Chien, K. Y., Chiang, C. M., Hseu, Y. C., Vyas, A. A., Rule, G. S., and Wu, W., 1994, *J Biol Chem*, 269(20), 14473-14483
131. Efremov, R. G., Volynsky, P. E., Nolde, D. E., Dubovskii, P. V., and Arseniev, A. S., 2002, *Biophys J*, 83(1), 144-153
132. Dubovskii, P. V., Lesovoy, D. M., Dubinnyi, M. A., Konshina, A. G., Utkin, Y. N., Efremov, R. G., and Arseniev, A. S., 2005, *Biochem J*, 387(Pt 3), 807-815
133. Forouhar, F., Huang, W. N., Liu, J. H., Chien, K. Y., Wu, W. G., and Hsiao, C. D., 2003, *J Biol Chem*, 278(24), 21980-21988
134. Wu, P. L., Lee, S. C., Chuang, C. C., Mori, S., Akakura, N., Wu, W. G., and Takada, Y., 2006, *J Biol Chem*, 281(12), 7937-7945
135. Rahman, S., Flynn, G., Aitken, A., Patel, Y., Hussain, F., Lu, X., Loftus, J. C., French, D., Wijelath, E., Strand, K., and Savidge, G. F., 2000, *Biochem J*, 345 Pt 3, 701-709
136. Wattam, B., Shang, D., Rahman, S., Egglezou, S., Scully, M., Kakkar, V., and Lu, X., 2001, *Biochem J*, 356(Pt 1), 11-17
137. Karlsson, E., Mbugua, P. M., and Rodriguez-Ithurralde, D., 1984, *J Physiol (Paris)*, 79(4), 232-240
138. Harel, M., Kleywegt, G. J., Ravelli, R. B., Silman, I., and Sussman, J. L., 1995, *Structure*, 3(12), 1355-1366
139. Bourne, Y., Taylor, P., and Marchot, P., 1995, *Cell*, 83(3), 503-512
140. le Du, M. H., Housset, D., Marchot, P., Bougis, P. E., Navaza, J., and Fontecilla-Camps, J. C., 1996, *Acta Crystallogr D Biol Crystallogr*, 52(Pt 1), 87-92
141. Bui, J. M., and McCammon, J. A., 2006, *Proc Natl Acad Sci U S A*, 103(42), 15451-15456
142. de Weille, J. R., Schweitz, H., Maes, P., Tartar, A., and Lazdunski, M., 1991, *Proc Natl Acad Sci U S A*, 88(6), 2437-2440
143. Yasuda, O., Morimoto, S., Jiang, B., Kuroda, H., Kimura, T., Sakakibara, S., Fukuo, K., Chen, S., Tamatani, M., and Ogihara, T., 1994, *Artery*, 21(5), 287-302
144. Banerjee, Y., Mizuguchi, J., Iwanaga, S., and Kini, R. M., 2005, *J Biol Chem*, 280(52), 42601-42611
145. Chuman, Y., Nobuhisa, I., Ogawa, T., Deshimaru, M., Chijiwa, T., Tan, N. H., Fukumaki, Y., Shimohigashi, Y., Ducancel, F., Boulain, J. C., Menez, A., and Ohno, M., 2000, *Toxicon*, 38(3), 449-462
146. Svetlov, M. S., Kommer, A., Kolb, V. A., and Spirin, A. S., 2005, *Protein Sci*,
147. Le Du, M. H., Ricciardi, A., Khayati, M., Menez, R., Boulain, J. C., Menez, A., and Ducancel, F., 2000, *J Mol Biol*, 296(4), 1017-1026
148. Ricciardi, A., le Du, M. H., Khayati, M., Dajas, F., Boulain, J. C., Menez, A., and Ducancel, F., 2000, *J Biol Chem*, 275(24), 18302-18310
149. Magis, C., Gasparini, D., Lecoq, A., Le Du, M. H., Stura, E., Charbonnier, J. B., Mourier, G., Boulain, J. C., Pardo, L., Caruana, A., Joly, A., Lefranc, M., Masella, M., Menez, A., and Cuniasse, P., 2006, *J Am Chem Soc*, 128(50), 16190-16205
150. Stricher, F., Martin, L., and Vita, C., 2006, *Methods Mol Biol*, 340, 113-149





---

## **ANNEXE 2**

---

Article : « Conserved structural determinants in three-fingered protein domain »

Andrzej Galat, Grégory Gross, Pascal Drevet, Atsushi Sato and André Ménez. (2008) FEBS J., 275, 3207-25.



# Conserved structural determinants in three-fingered protein domains

Andrzej Galat<sup>1</sup>, Gregory Gross<sup>2</sup>, Pascal Drevet<sup>2</sup>, Atsushi Sato<sup>3</sup> and André Ménez<sup>4,\*</sup>

<sup>1</sup> Institut de Biologie et de Technologies de Saclay, SIMOPRO/DSV/CEA, Gif-sur-Yvette, France

<sup>2</sup> Institut de Biologie et de Technologies de Saclay, SBIGeM/DSV/CEA, Gif-sur-Yvette, France

<sup>3</sup> Department of Information Science, Faculty of Liberal Arts, Tohoku-Gakuin University, Sendai, Japan

<sup>4</sup> Muséum National d'Histoire Naturelle, Paris, France

## Keywords

atomic interactions; cystine networks; three-finger proteins; three-fingered protein; three-fingered protein domain

## Correspondence

A. Galat, Bat. 152, CE-Saclay, F-91191

Gif-sur-Yvette Cedex, France

Fax: +33 1 69 08 90 71

Tel: +33 1 69 08 84 67

E-mail: galat@dsvidf.cea.fr

\*Deceased. The former President of the Museum of Natural History, Paris, France

(Received 6 March 2008, revised 17 April 2008, accepted 18 April 2008)

doi:10.1111/j.1742-4658.2008.06473.x

The three-dimensional structures of some components of snake venoms forming so-called 'three-fingered protein' domains (TFPDs) are similar to those of the ectodomains of activin, bone morphogenetic protein and transforming growth factor- $\beta$  receptors, and to a variety of proteins encoded by the *Ly6* and *Plaur* genes. The analysis of sequences of diverse snake toxins, various ectodomains of the receptors that bind activin and other cytokines, and numerous gene products encoded by the *Ly6* and *Plaur* families of genes has revealed that they differ considerably from each other. The sequences of TFPDs may consist of up to six disulfide bonds, three of which have the same highly conserved topology. These three disulfide bridges and an asparagine residue in the C-terminal part of TFPDs are essential for the TFPD-like fold. Analyses of the three-dimensional structures of diverse TFPDs have revealed that the three highly conserved disulfides impose a major stabilizing contribution to the TFPD-like fold, in both TFPDs contained in some snake venoms and ectodomains of several cellular receptors, whereas the three remaining disulfide bonds impose specific geometrical constraints in the three fingers of some TFPDs.

To date, more than 45 000 protein three-dimensional structures have been deposited in the Protein Data Bank (PDB) [1], many of which have a high sequence similarity to each other. Analyses of these structures have revealed approximately 1000 diverse polypeptide chain folds [2], as predicted about 10 years ago [3]. This number, however, may be subject to debate because of the various possible ways of defining protein folds [4,5]. Nevertheless, it is accepted that the space of protein folds is considerably smaller than that of protein sequences [6,7]. However, how a given protein fold may evolve towards a novel function remains obscure [6,7]. One way to approach such a complex question is to analyse a set of functionally different

proteins recognized to adapt the same fold, and to search for structural determinants that may reflect both divergence and convergence criteria that are critical to the fold [5–9].

This study aims to identify the determinants associated with the three-dimensional structure of a fold that characterizes a group of homologous proteins rich in disulfides. According to the SCOP server (<http://scop.mrc-lmb.cam.ac.uk/scop>) [2], approximately 75 folds are considered to be relatively small in size, and about 50 are rich in disulfide bonds. In this study, we focused our work on a group of proteins adapting the fold originally discovered for snake neurotoxins, which possesses three adjacent fingers rich in  $\beta$ -pleated sheets

## Abbreviations

Act-R, activin receptor; BMP-R, bone morphogenetic protein receptor; ECD, ectodomain; GPCR, G-protein-coupled receptor; ID, sequence similarity score; MSA, multiple sequence alignment; TFP, three-fingered protein; TFPD, three-fingered protein domain; TGF $\beta$ -R, transforming growth factor- $\beta$  receptor; TM, transmembrane segment; uPAR, urokinase/plasminogen activator receptor; WGA, wheatgerm agglutinin.

[10–12]. In order to provide proteins of this group with a historically accepted name and a relevant topographical designation, we have called them three-fingered proteins (TFPs), which all share one or more three-fingered protein domains (TFPDs). In this article, we describe the analyses of fifty three-dimensional structures of diverse TFPDs [1] and several hundreds of sequences containing the TFPD-like motif.

A TFPD possesses the following features. Firstly, it is made up of a single polypeptide chain of 60–100 amino acid residues, folded into three adjacent loops emerging from a hydrophobic palm, which includes at least three and, in the majority of cases, four disulfide bonds. Secondly, it possesses five  $\beta$ -strands encompassing the three loops or fingers. Thirdly, the TFPDs act as monomers or multimers, and display substantial variations in terms of loop size and shape, number of extra disulfide bonds and additional secondary structures. Fourthly, the TFPDs display a wide distribution in the eukaryotic kingdom. Fifthly, the TFPDs are devoid of known enzymatic activities, but exert a wide range of binding activities, varying from ligands (including toxins that block or modulate the functions of different receptors, ion channels and enzymes [13]) to receptors that are anchored to the cell surface membrane [such as CD59 or urokinase/plasminogen activator receptor (uPAR), also known as CD87]. Activin (Act-R), bone morphogenetic protein (BMP-R) and transforming growth factor- $\beta$  (TGF $\beta$ -R) receptors [14] transmit signals through a transmembrane (TM) segment to their cytoplasmic kinase domains.

Cheek *et al.* [15] have recently classified small proteins rich in disulfide bonds into 41 different fold groups. Three of these are called 'knottin-like I, II and III', which are characterized by a structural core consisting of four cysteine residues forming a disulfide crossover. According to these authors, the TFPDs belong to 'knottin-like group II'. Interestingly, despite the fact that some plant lectins, such as wheatgerm agglutinin (WGA), are considered to share some topographical similarity with TFPDs [16], they have been classified to a different fold, namely 'knottin-like group I'. According to Cheek *et al.* [15], the four cysteines are located on four elements that adapt different spatial connections in groups I and II. In this work, we have analysed in detail the conserved structural elements of the TFPDs and examined whether or not they are also present in some plant lectins.

We have found that all analysed TFPDs share a conserved structural core that includes two small  $\beta$ -sheets encompassing the three loops (fingers), a network of three cysteines and several clusters of inter-atomic interactions, including one cluster that involves

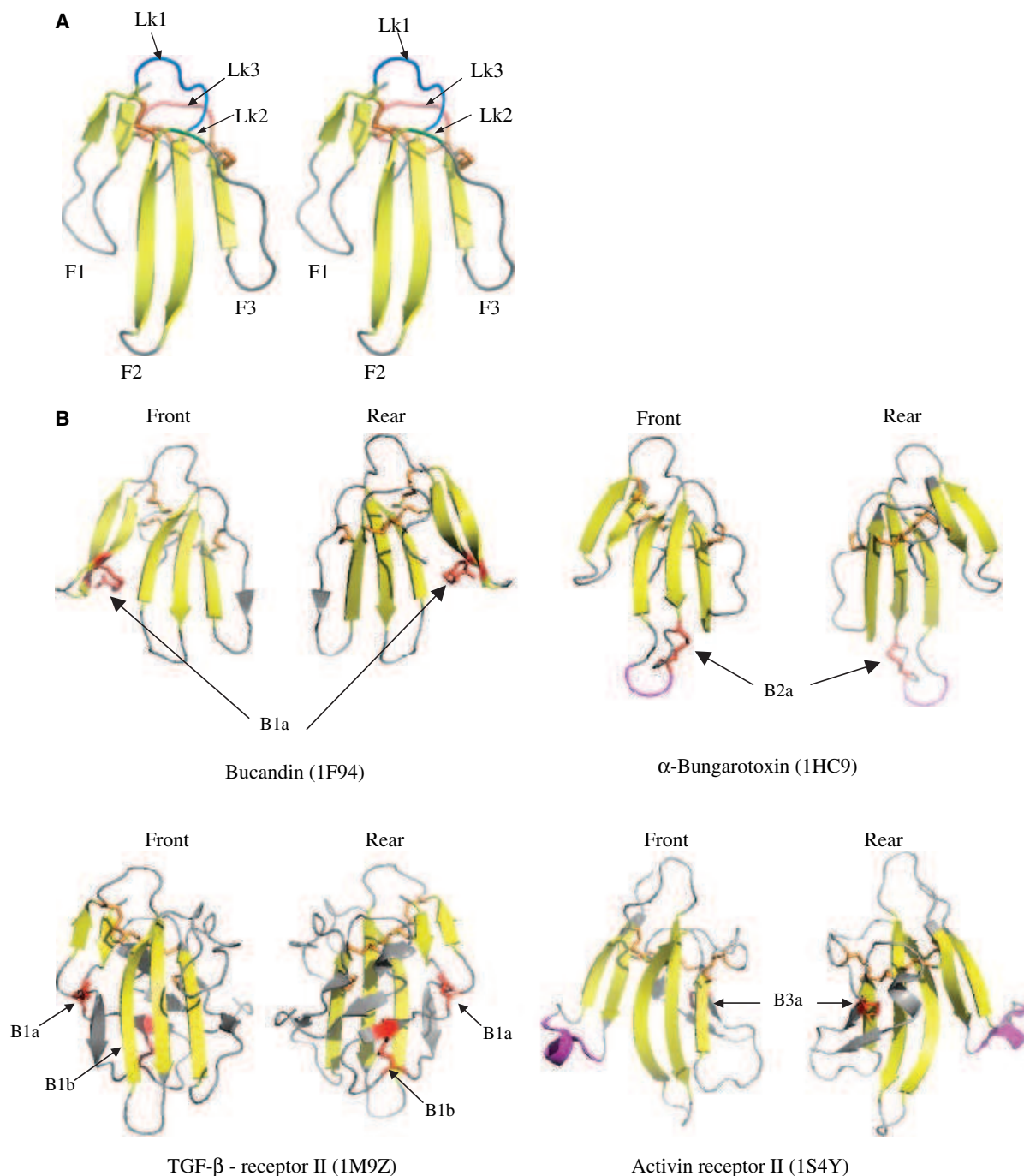
a strictly conserved asparagine residue, which establishes several hydrogen bonds with the amino acids in the three fingers. We have accumulated evidence suggesting that the cystine that locks the third finger is differently organized in the TFPDs that act as ligands or receptors. Finally, our definition of the TFPD fold has allowed for its clear distinction from the fold typical of several plant lectins, such as WGA.

## Results and Discussion

### On the diversity of TFPDs

In Fig. 1, the three-dimensional structure (1IQ9) of a typical TFP, i.e. a short-chain neurotoxin from snake venom, is shown. The four disulfide bonds form a tight network at the base of a palm, from which emerge three long loops, called fingers F1, F2 and F3. A disulfide bridge tightly closes each finger. F1 is linked to F2 and F2 to F3 by  $\beta$ -turns called Lk1 and Lk2, respectively. The Lk3 turn includes four amino acid residues forming a  $\beta$ -turn closed by the last disulfide bridge of the molecule. The  $\beta$ -sheet in F1 includes two  $\beta$ -strands ( $\beta$ 1– $\beta$ 2) linked by a  $\beta$ -turn at the tip of F1, whereas the second small  $\beta$ -sheet involves three  $\beta$ -strands ( $\beta$ 3– $\beta$ 4– $\beta$ 5) located on F2 and F3. The three fingers point approximately in the same direction.

In Table 1, data are summarized on the TFPDs whose three-dimensional structures have been used in this work. The 34 selected toxins from snake venoms act as blockers or modulators of ligand-gated ion channels (snake neurotoxins), integrin receptors (dendroaspin), enzymes (fasciculins) or G-protein-coupled receptors (GPCRs) interacting with muscarinic toxins. Table 1 also includes 16 structures of cell surface membrane-bound proteins, such as uPAR, Act-R and TGF $\beta$ -R. NIR represents the number of intramolecular atomic interactions calculated in the range 2.7–4.5 Å (2.7–4.0 Å). NIR is the sum of the intramolecular interactions whose nature varies with the overall hydrophobicity of a given TFPD. There are about 28–31% interactions between diverse C and S atoms (hydrophobic interactions) and 15–18% interactions between diverse O and N atoms (hydrophilic interactions); the remainder is caused by interactions between the atoms from these two groups. Although, the spatial organizations of some secondary structures in the diverse TFPDs are similar, the distributions of the atomic interactions vary. Thus, about 32–34% interactions occur between atoms in the main chain, 22–31% between atoms of diverse side chains and the remainder between main chain atoms and side chain atoms.



**Fig. 1.** (A) Stereoview of the tertiary structure of a TFP: the  $\alpha$ -neurotoxin of *Naja nigricollis* (1IQ9). The structure was annotated as follows: F1, F2 and F3 indicate the three successive fingers and Lk1, Lk2 and Lk3 denote the linkers that join F1 to F2, F2 to F3 and F3 to the C-terminal, respectively. (B) Front and rear views of spatial positioning of the disulfides B1a, B2a, B2b and B3a.

The length of the polypeptide chain of a TFPD may vary from 59 to 106 amino acids, except for uPAR which contains three consecutive TFPDs. The number of interatomic interactions shorter than 4.5 Å varies from about 1100 pairs for an average sized short

neurotoxin structure to almost twice as many in the larger ectodomain (ECD) of TGF $\beta$ -RII. Obviously, this number depends on several factors, including the structural resolution. In this respect, NMR-based structures must be considered with caution.

**Table 1.** Crystallographic structures of diverse TFPDs. Ab, antibody; NIR, number of intramolecular atomic interactions below 4.5 Å (4 Å); Norm-B factors show the most flexible parts of the molecule (calculated for the C $\alpha$  atoms); NR, number of amino acids used in the analysis.

No.	PDB	Protein (complex)	Organism	R (Å)	NR	NIR/4.5 Å (4 Å)	Norm-B	Reference
Toxins from diverse snake venoms								
T1	1IQ9	Toxin a	<i>Naja nigricollis</i>	1.80	61	1128 (521)	18P, 19G, 48G	[17]
T2	1VBO	Atratoxin-B	<i>N. atra</i>	0.92	61	1150 (575)	19G, 33G	[18]
T3	1JE9	Neurotoxin II	<i>N. kaouthia</i>	NMR	61	964 (472)		[19]
T4	2ERA	Erabutoxin A, S8G	<i>Laticauda semifasciata</i>	1.80	62	1116 (536)	45TVK47	[20]
T5	1QKE	Erabutoxin A	<i>L. semifasciata</i>	1.50	62	1103 (532)	10E, 45TVK47	[21]
T6	6EBX	Erabutoxin B	<i>L. semifasciata</i>	1.70	62	1142 (552)	20G, 47KPG49	[22]
T7	1FAS	Fasciculin-I	<i>Dendroaspis angusticeps</i>	1.80	61	1074 (498)	7TTTSRAI13	[23]
T8	1FSC	Fasciculin-II	<i>D. angusticeps</i>	2.00	61	1083 (503)	19G, 32K, 33M, 55S	[24]
T9	1FSS	Fasciculin-II/(AChE)	<i>D. angusticeps</i>	1.90	61	1097 (513)	18GE19, 43P, 44G, 54T	[25]
T10	1F8U	Fasciculin-II/(AChM)	<i>D. angusticeps</i>	2.90	61	1082 (543)	18GEN20, S55	[26]
T11	1FF4	Muscarinic toxin 2	<i>D. angusticeps</i>	1.50	65	1248 (562)	7KSIGG11	[27]
T12	1F94	Bucandin	<i>Bungarus candidus</i>	0.97	63	1267 (610)	19AE20, 22T, 42T, 44TE45	[28]
T13	2H8U	Bucain	<i>B. candidus</i>	2.20	65	1022 (468)	32NPSGK	[29]
T14	1JGK	Candoxin	<i>B. candidus</i>	NMR	66	1027 (478)		[30]
T15	2H5F	Denmotoxin	<i>B. dendrophila</i>	1.90	75	1225 (581)	41DENG45	[31]
T16	2H7Z	Iridotoxin	<i>B. dendrophila</i>	1.50	75	1302 (578)	17TSSDCS	[31]
T17	1TGX	Cardiotoxin	<i>N. nigricollis</i>	1.55	60	878 (373)	16K, 28A, 32V, 33P	[32]
T18	1CXO	Cardiotoxin	<i>N. nigricollis</i>	NMR	60	1285 (643)		[33]
T19	1H0J	Cardiotoxin-3	<i>N. atra</i>	1.90	60	1083 (492)	12K, 16A, 17G, 23K, 24M, 49V	[34]
T20	2BHI	Cardiotoxin A3/ sulfogalactoceramide	<i>N. atra</i>	2.31	60	1047 (486)	8PLF, 22Y, 31KV	[35]
T21	1UG4	Cardiotoxin-IV	<i>N. atra</i>	1.60	60	1033 (502)	28AAPLVP33	[36]
T22	1CDT	Cardiotoxin	<i>N. mossambica</i>	2.50	60	1059 (503)	29K	[37]
T23	1KXI	Cardiotoxin-V	<i>N. n. atra</i>	2.19	62	971 (438)	17E, 29K, 30F	[38]
T24	1CHV	Cardiotoxin-(analogue)	<i>N. n. atra</i>	NMR	60	874 (415)		[39]
T25	1CB9	Cardiotoxin	<i>N. oxiana</i>	NMR	60	823 (380)		[40]
T26	2CTX	$\alpha$ -Cobratoxin	<i>N. n. siamensis</i>	2.40	71	1121 (510)	67-TRKRP-71	[41]
T27	1LXG	$\alpha$ -Cobratoxin/ (YRGWKHWVYYTCCPDTPYLhS)	<i>N. n. kaouthia</i>	NMR	71	998 (515)		[42]
T28	1YI5	$\alpha$ -Cobratoxin/acetylcholine binding protein (AChB)	<i>N. n. siamensis</i>	4.20	68	907 (396)		[43]
T29	1HC9	$\alpha$ -Bungarotoxin/ (WRYYESLLPYPD)	<i>B. multicinctus</i>	1.80	74	1296 (551)	50SKKPY54, C-term	[44]
T30	1NTN	Neurotoxin-I	<i>N. n. oxiana</i>	1.90	72	1110 (524)	C-term	[45]
T31	1KBA	$\kappa$ -Bungarotoxin	<i>B. multicinctus</i>	2.30	66	1222 (583)	15P, 16N, 17G, 35G	[46]
T32	1KFH	$\alpha$ -Bungarotoxin	<i>B. multicinctus</i>	NMR	74	1612 (836)		[47]
T33	1LSI	Long neurotoxin	<i>L. semifasciata</i>	NMR	66	1162 (569)		[48]
T34	1DRS	Dendroaspin	<i>D. j. kaimose</i>	NMR	59	923 (443)		[49]
Ectodomains of some receptors								
R1	1CDR	CD59/(disaccharide)	<i>Homo sapiens</i>	NMR	77	1256 (569)		[50]
R2	2OFS	CD59	<i>H. sapiens</i>	2.12	75	1512 (684)	32GLQ	[51]
R3	1YWH	Urokinase receptor/ (KSDChaFskYLWSSK)	<i>H. sapiens</i>	2.70	268	4527 (1914)	79GNSGG, C-term	[52]
R4	2FD6	uPAR/plasminogen/Ab	<i>H. sapiens</i>	1.90	248	4642 (2091)	92L, 116SPPEE, 229EPKNQSY	[53]

**Table 1.** Continued.

No.	PDB	Protein (complex)	Organism	R (Å)	NR	NIR/4.5 Å (4 Å)	Norm-B	Reference
R5	2I9B	uPAR/plasminogen	<i>H. sapiens</i>	2.60	265	4414 (1957)		[54]
R6	1BTE	Act-RIIA	<i>Musculus musculus</i>	1.50	97	1944 (787)	33G, 38R, 61LDDIN65	[56]
R7	1LX5	Act-RIIA/(BMP7)	<i>H. sapiens</i>	3.30	94	1304 (913)		[56]
R8	1S4Y	Act-RIIB/(Inhibinβa)	<i>M. musculus</i>	2.30	91	1723 (790)	29GEQD32	[57]
R9	1NYU	Act-RIIB/(Inhibinβa)	<i>Rattus norvegicus</i>	3.10	92	1699 (760)	26T, 50EGE52, 67SG68	[58]
R10	2HLR	BMP-RII	<i>Ovis aries</i>	1.20	67	626 (434)	39PY, 78N	[59]
R11	1REW	BMP-RIA/(BMP2)	<i>H. sapiens</i>	1.86	89	1457 (677)	47DAIN50, 67DQ68, 109QYLQ112	[60]
R12	1ES7	(BMP-RAI) <sub>2</sub> /(BMP2)	<i>H. sapiens</i>	2.90	83	1304 (585)	265ED266, 270270	[61]
R13	2H64	Act-RIIB/BMIRA/BMP2	<i>H. sapiens</i> / <i>M. musculus</i> / <i>H. sapiens</i>	1.92	92	1476 (700)	67DQ	[62]
R14	2GOO	Act-RIIA/BMIRA/BMP2	<i>H. sapiens</i> / <i>M. musculus</i> / <i>H. sapiens</i>	2.20	92	1860 (662)	60WL	[63]
R15	1M9Z	TGFβ-RII	<i>H. sapiens</i>	1.05	105	2030 (951)	104KKPG107, C-term	[64]
R16	1KTZ	TGFβ-RII/(TGFβ3)	<i>H. sapiens</i>	2.15	106	2064 (949)	25P, 91E	[65]

### Conserved and variable sequence features of TFPDs

In Fig. 2, an alignment of the non-redundant primary structures of the three-fingered ligands and ECDs listed in Table 1 is shown. Using the sequence of the short neurotoxin from *Naja nigricollis* (1IQ9) as an arbitrary reference, we calculated the pairwise sequence similarity scores (IDs) with the remaining sequences of the other TFPDs (Fig. 2), and found that they varied between 86% and 30% for diverse snake toxins and below 25% for the ECD sequences of some cell surface receptors. This difference is caused, at least in part, by the longer loops of the ECDs and extensive amino acid substitutions in the fingers. In Fig. 2, a number of strictly conserved sequence features are emphasized. These include six half-cystines that form three disulfides, named B1, B2 and B4, five β-strands (coloured yellow) located on fingers 1, 2 and 3, and an asparagine residue adjacent to the last half-cystine of B4. These are the minimal strictly conserved sequence and structural features that define the TFPD based on the alignment of sequences from the three-dimensional structures.

Other sequence features are highly but not strictly conserved. These include the cystine called B3, which is only lacking in the first domain of uPAR (1YWH1), a hydrophobic residue (often an aromatic residue) adjacent downstream to the second half-cystine of B1, and a glycine residue adjacent upstream to the second

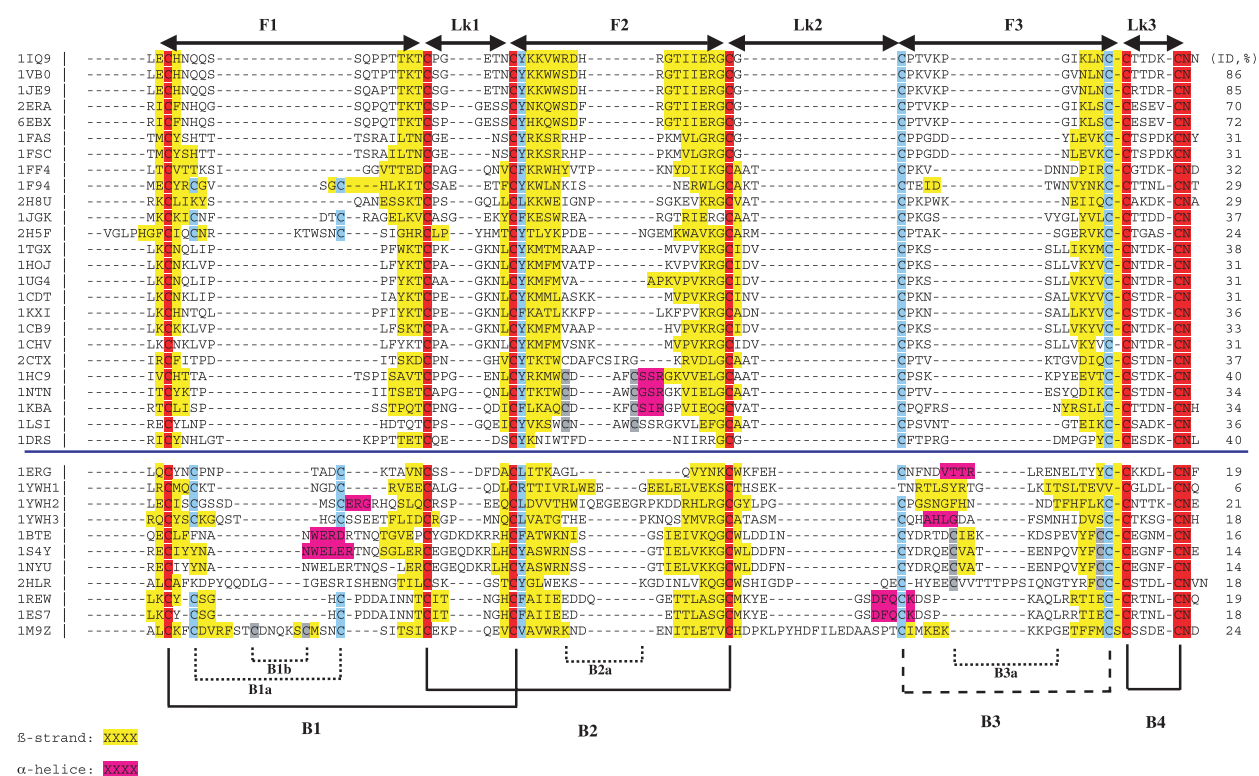
half-cystine of B2. This glycine residue is strictly conserved in all the toxins only. In addition, linker 1 usually comprises four to six amino acids, except for several ECDs where it can be as long as nine amino acids (ActRIIb). Similarly, linker 3 comprises four amino acids, except in two cases where it can be five amino acids (fasciculin). Other sequence elements of TFPD tend to vary substantially from one protein to another. These include the length and composition of the fingers, small helical stretches and additional disulfides, which are labelled by a letter related to the disulfide that surrounds them (Fig. 2). With the exception of B1a, the disulfide bridges seem to be specific to certain classes of TFPD (Fig. 2), such as B2a which occurs in long neurotoxins and B3a which is found in Act-RII. B1a is a more common feature and can be seen in both ligands, such as bucandin, and in the ECDs of receptors (e.g. TGFβ-R); in contrast, B1b only occurs in the ECDs of TGFβ-RII (Fig. 1B).

### On the conserved and variable three-dimensional features of TFPDs

#### Conserved interaction clusters

To compare qualitatively and quantitatively the three-dimensional structures of diverse TFPDs, distance maps were constructed from the three-dimensional structures (Table 1). Figure 3 illustrates such maps calculated for two three-fingered ligands and two three-fingered ECDs. Figure 3A shows a comparison





**Fig. 2.** Alignment of unique sequences from the structures listed in Table 1. The optimal alignment of half-cystines was obtained by introducing a few gaps manually. The amino acids in the  $\beta$ -sheet and  $\alpha$ -helical structures are shown in yellow and magenta, respectively. Strictly conserved amino acids are shown in red, highly conserved half-cystines in blue and class-specific half-cystines in grey. Arrows at the top of the aligned sequences encompass amino acids belonging to fingers 1, 2 and 3 (F1, F2, F3) and to linkers Lk1, Lk2 and Lk3. Disulfide bridges were named as B1, B2, etc., as indicated.

between the distance maps of the  $\alpha$ -neurotoxin from *N. nigricollis* (1IQ9, bottom triangle on left of diagonal) and the ECD of Act-RIIB bound to Act (1S4Y, top triangle on right of diagonal) [57]. Figure 3B shows the distance maps of  $\alpha$ -bungarotoxin (1HC9, bottom triangle) and the third TFPD of uPAR (1YWH, top triangle).

We made a similar two-by-two comparison for all the TFPDs shown in Table 1, and found that all display similar distributions of common interaction clusters. Thus, three readily recognizable main clusters are associated with the three fingers. They correspond to interactions between  $\beta$ 1 and  $\beta$ 2 (cF1, coloured pink),  $\beta$ 3 and  $\beta$ 4 (cF2, coloured blue) and  $\beta$ 5 with the extended loop linking  $\beta$ 4 to  $\beta$ 5 (cF3, coloured pink). Conserved clusters are also observed at the interfaces [indicated as (i)] between the fingers (iF1/F2 and iF2/F3) and between finger 1 and linker 1 (cF1/Lk1). In addition, a super-cluster of interactions involving three smaller clusters [Lk3/ $\beta$ (1), Lk3/ $\beta$ (3), Lk3/ $\beta$ (4), coloured violet] is seen between the C-terminal  $\beta$ -turn and three  $\beta$ -strands. In total, nine homologous clusters

(coloured ellipses) were found in all TFPDs, together with some scattered small islands of atomic interactions that often implicate disulfide bridges (indicated as B and shown by red squares). These nine clusters form a conserved structural core in all the analysed TFPDs.

However, the relatively large differences in the lengths of the polypeptide chains of the TFPDs sometimes introduce additional secondary structures to the minimal TFP fold represented by the structures of short neurotoxins, such as erabutoxins A and B [10–12]. As a result, some differences in the interaction patterns were detected in several distance maps. Thus, finger F3 is longer in the ECDs of the receptors in comparison with the toxins. This is particularly well illustrated on the distance map of the ECD of Act-RIIB (1S4Y, Fig. 3A). Its finger cF3 possesses two additional  $\beta$ -strands ( $\beta$ 4a and  $\beta$ 5), which establish strong interactions with each other (see the large pink-coloured cluster in the bottom part of the right side of Fig. 3A). In addition, F1 not only includes  $\beta$ 1 and  $\beta$ 2, like the other TFPDs, but also a short  $\alpha$ -helix and a

$\beta$ -turn. Finally, the additional  $\beta$ -strand ( $\beta 6$ ), which is the last secondary structure before the TM segment that links the ECD of Act-RIIB with an intracellular kinase domain, interacts with  $\beta 3$ ,  $\beta 4a$  and a tyrosine residue in  $\beta 5$ . The  $\beta$ -strands are longer in the third domain of uPAR and are spaced by longer runs of  $\beta$ -turns and  $\alpha$ -helices. Similar networks of atomic interactions were observed in the distance maps of the two other domains of uPAR (data not shown). A distance map of the entire uPAR (data not shown) indicated that, in addition to the atomic interactions inherent to each of the three TFPDs, some atomic interactions can also be seen between domains I, II and III.

### Deeper analysis of the interaction clusters

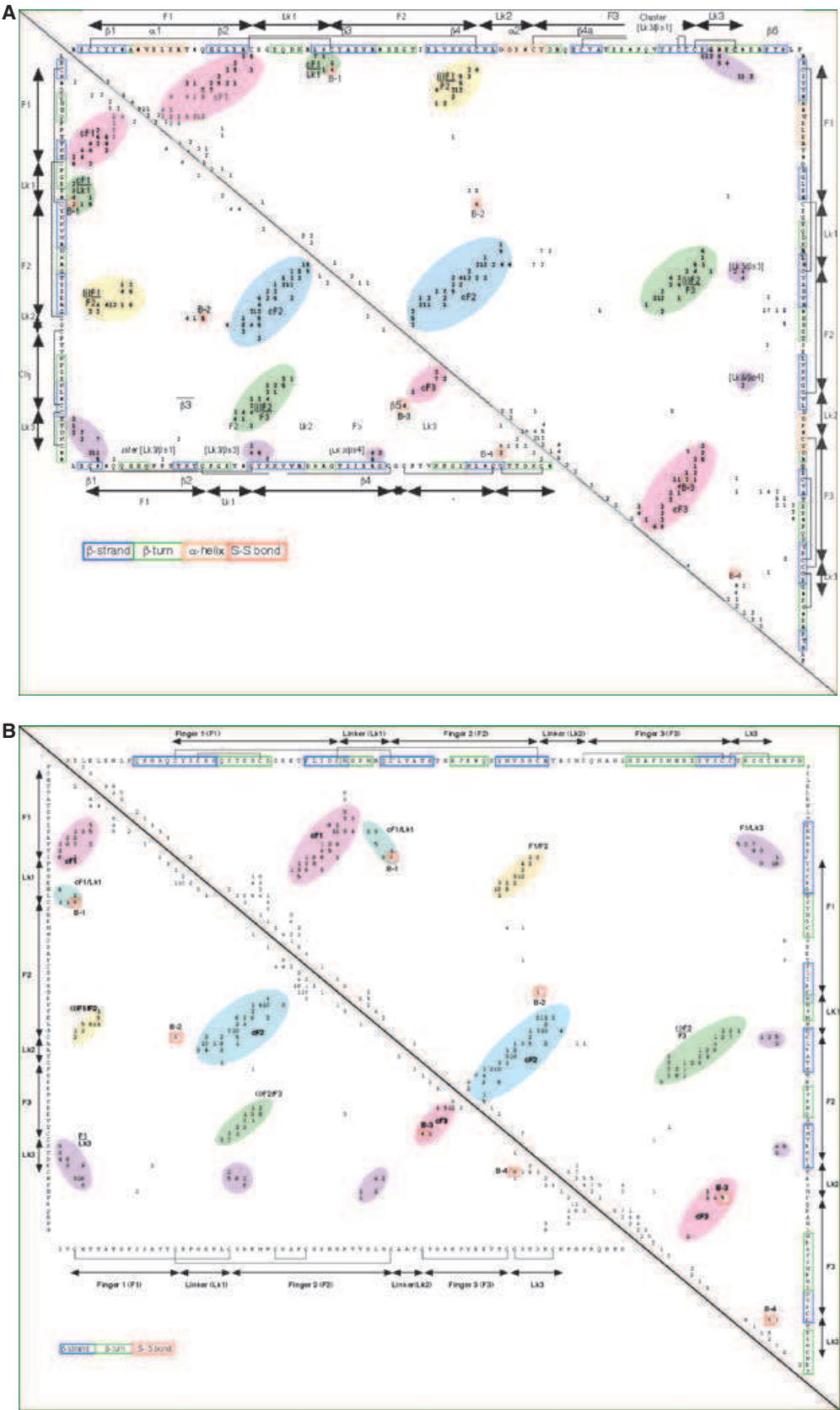
Using distance matrices, specific intramolecular interaction networks and calculated levels of their conservation, we established the variations of these three measures in the different TFPDs shown in Table 1. For example, in order to further document the intramolecular interaction networks for the  $\alpha$ -toxin of *N. nigricollis* (1IQ9, Fig. 3A, bottom panel) and the third TFPD of human uPAR (1YWH3, Fig. 3B, top panel), we summed the numbers of distances below 4.5 Å for each amino acid residue and calculated their non-bonding van der Waals' and Coulombic interactions. The diagrams in Fig. 4A, B show the number of distances scaled down by a factor of 0.1 (top panel) and the sum of the van der Waals' and Coulombic energy terms (bottom panel) for the atomic interactions within these two TFPDs (for  $d \leq 4.5$  Å). These linear diagrams show that several of the amino acids establish higher than average numbers of interactions and, consequently, become the main contributors to the overall stability of the TFPDs. For example, the data shown in Fig. 4B reveal that 37 amino acids of the third TFPD of human uPAR (1YWH3) establish more than 20 contacts, whereas no more than 13 amino acids establish more than 30 contacts. About 15 amino acids are seen to establish a large proportion of van der Waals' and electrostatic interactions.

The data shown in Fig. 4A,B are typical of that seen for all the remaining TFPDs. In all cases, the largest number of contacts and the best energy terms are attributed to the half-cystines, and to several amino acids in their vicinity. More precisely, in supplementary Table S1, the numbers of interactions established by B1, B2, B3 and B4 and some of their neighbouring amino acids, including the conserved asparagine that is adjacent to the second half-cystine of B4, are listed. Some general trends emerge from the data shown in

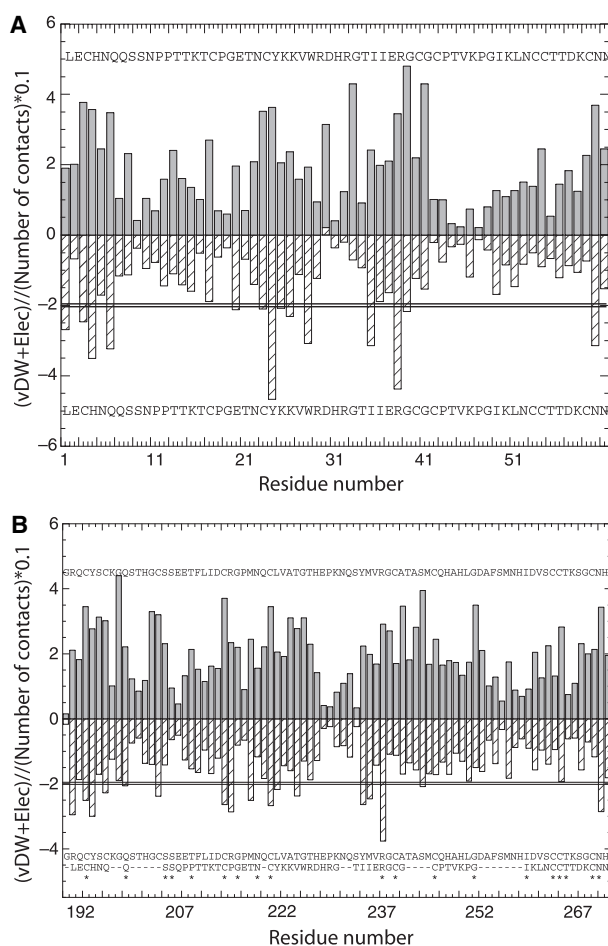
supplementary Table S1. Thus, a particularly large number of contacts can be observed for the half-cystines C1, C3 and C4, together with some of their neighbouring amino acids. This is particularly obvious for C3 and the conserved hydrophobic residue that follows it (often an aromatic residue), and for C4 and its preceding conserved adjacent sequence (often RG in toxins). These two half-cystines and their conserved neighbours seem to be crucial stabilizing factors in TFPDs, especially in the toxins. In a few cases, the numbers of interactions on the C-terminal aspartic acid can be substantially lower, as for 1LSI, whose NMR-established structures show, on average, only 11 atomic distances below 4.5 Å. This is also the case for the ECD of TGF $\beta$ -RIIB but, in this example, the amino acids following the CN doublet have a large number of interactions as they link the TFPD to the TM segment. In addition, in dendroaspasin (1DRS), the asparagine establishes a small number of contacts below 4.5 Å; however, the leucine residue that follows the CN doublet displays a large number of contacts below 4.5 Å. B3 and, especially, its first half-cystine C5 establish a smaller number of contacts and a smaller energy contribution than the three other strictly conserved S-S bonds B1, B2 and B4, suggesting that B3 is less crucial in the maintenance of the TFPD structure, a view which agrees with the observation that this bond is lacking in TFPD-I of uPAR (1YWH.1 in supplementary Table S1). The energy contributions of the fifth S-S bond B2a (e.g. bucandin or long neurotoxins) and the sixth S-S bond B1b (ECD of TGF $\beta$ -RIIB) are comparable with those of the three bonds B1, B2 and B4 (data not shown).

Therefore, the histograms illustrated in Fig. 4 demonstrate that the strictly conserved cystines B1, B2 and B4 and some adjacent amino acids show both a large number of atomic contacts and important energy contributions, suggesting that these amino acids are crucial for the stability of TFPDs. Our data also show, however, that some individual amino acids with a high conservation level in some groups of TFPDs do not necessarily have similar contributions to the stability of each TFPD. For example, the hydrophobic amino acid residue that follows the second half-cystine of B1 [see supplementary material for the multiple sequence alignment (MSA) of diverse TFPDs] does not establish a similar number of atomic contacts and energy contributions in the toxin and TFPD-III of uPAR.

The strictly conserved asparagine that is adjacent to C8 (the highly conserved CN sequence motif) is also involved in a large number of interactions (supplementary Table S1). Its side chain is oriented towards the

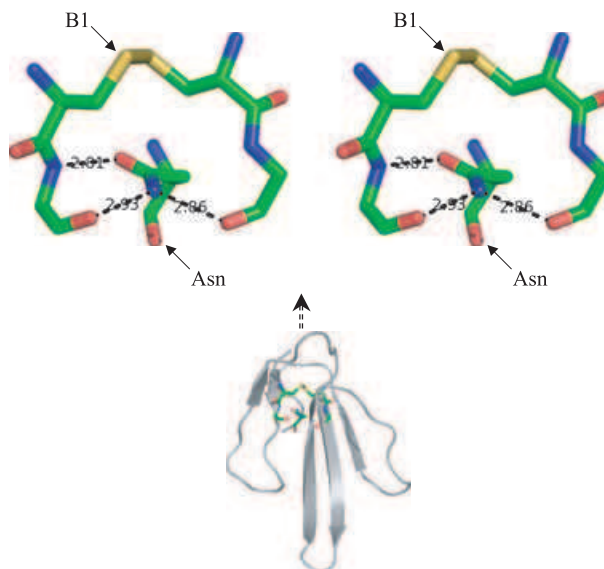


**Fig. 3.** Bi-triangular distance maps of four TFPDs. (A) ECD of Act-RIIB (1S4Y, top triangle) and the short neurotoxin from *Naja nigricollis* (1IQ9, bottom triangle); (B) TFPD-III from human uPAR (1YWU, top triangle) and  $\alpha$ -bungarotoxin (1HC9, bottom triangle). The amino acid sequence of each protein is shown vertically and horizontally on one side of the diagonal. The clusters of intramolecular interactions equal to or below 4 Å are indicated by coloured ovals. The red squares correspond to disulfides B1–B4.



**Fig. 4.** All the atomic contacts per amino acid residue scaled down by a factor of 0.1 (top panels) and sequence distribution of the sum of van der Waals' (vdW) and Coulombic (Elec) terms (bottom panels): (A) TFPD of the  $\alpha$ -toxin of *Naja nigricollis* (1IQ9); (B) third TFPD of human uPAR (1YWH).

interior of all the TFPDs, as shown in Fig. 5, except in dendroaspasin where it points in the opposite direction. We suspect that this peculiar behaviour may be related to the low-resolution NMR structure of this toxin. As shown in supplementary Table S1, the atoms of the asparagine residue establish large numbers of atomic interaction pairs ( $\leq 4.5$  Å). We found that some of these interactions, at least one of the three shown in Fig. 5, are conservatively present in the different TFPDs. Thus, by interacting firmly with the upper part of F1 and F2, the side chain of the conserved

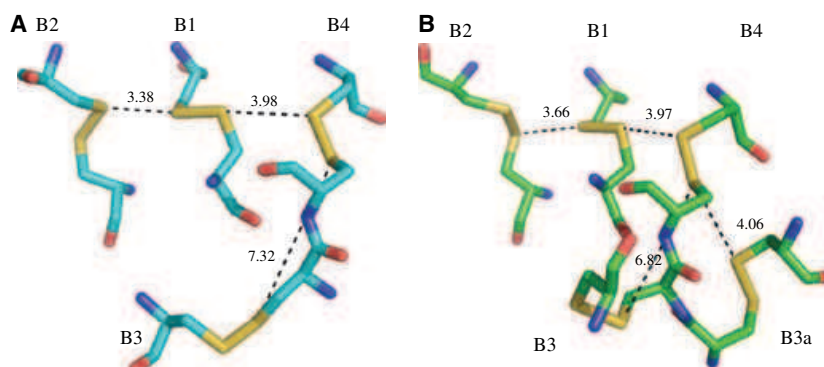


**Fig. 5.** Stereoview of the strictly conserved structural motif involving a loop formed by amino acids on the first and fourth  $\beta$ -strands linked by the disulfide bond B1, wrapped around the conserved asparagine (Asn) residue. Three conserved hydrogen bonds observed between the loop and the Asn residue are shown (1VB0).

asparagine locks the C-terminal part of the structure with two of the three fingers of the TFPD. In view of all these considerations, we propose that the assemblies involving B1, B2 and B4, some of their neighbouring amino acids and the C-terminal asparagine region constitute key stabilizing elements in all TFPDs.

### A structurally conserved cystine cluster

The most common type of cystine cluster is illustrated in Fig. 6A, which involves a tight clustering of the sulfur atoms in the disulfide pairs B1/B2 and B1/B4. Cysteine is an amino acid residue with a high hydrophobicity; in a recent study, it was assigned the highest hydrophobicity potential [67]. In the third finger of the ECD of Act-RIIB (1S4Y), B3A disulfide establishes a close contact with B4, as it is a part of the triplet of C-terminal cysteine residues (CCCxxxxCN assembly, see Fig. 6B). We also investigated the mode of stacking of the cystines using some of the concepts developed by Harrison and Steinberg [68]. Good stacking was observed in the majority of pairs B1/B2 and B1/B4, whereas for the majority of cases loose stacking was



**Fig. 6.** Cystine clusters in two TFPDs: (A) the  $\alpha$ -toxin of *Naja nigricollis* (1IQ9); (B) the ECD of mouse Act-RIIB (1S4Y). Made with the PYMOL program [66].

found for B1/B3, B2/B3 and B3/B4. There is no cross-over of any of these disulfides as seen from the top of the molecule, i.e. from the Lk1 direction.

Moreover, the two additional cystines, B1a and B1b, in the first finger of the ECD of TGF $\beta$ -RII (1KTZ) do not cluster with the remaining four cystines. All of these data support the idea that only the three conserved cystines B1, B2 and B4 form a strongly packed interaction network in the TFPD, whereas the other cystines are more or less apart from this tight network. The only exception is the interaction between B3 and B4 in the ECD of TGF $\beta$ -RII, but it is important to specify that the usually conserved doublet of the cysteine residues is split by an additional amino acid residue (see Fig. 2). Therefore, we called the B1/B2 and B1/B4 interaction network the 'conserved cystine cluster' [68].

To better characterize this cluster in all the TFPDs, we calculated the distances in the range  $\geq 3.0$  Å to  $\leq 7.5$  Å between the sulfur atoms of the cysteine residues, and the van der Waals' and Coulombic energy terms (interaction energy terms) for their interactions. Subtle variations of these values in the cystine clusters are shown in supplementary Fig. S1. In the majority of cases, the average S—S distance and interaction energies are clustered in a quasi-linear fashion, but several S—S networks have higher energy terms and come from the complexes of toxins bound to acetylcholine esterase, in which the interatomic distance in some of the S—S bonds is shorter than that in the free forms of the toxins. In the latter cases, some deformation of TFPD takes place on binding to the enzyme. In addition, we calculated the distances between the C $\alpha$  ( $\gamma^{\alpha}_{ij}$ ) and C $\beta$  ( $\gamma^{\beta}_{ij}$ ) atoms [69] in each cystine of the analysed TFPDs. In supplementary Fig. S2, the C $\beta$ —C $\beta$  distances are shown, which are clustered in the range 3.6–4.0 Å, whereas the C $\alpha$ —C $\alpha$  distances vary over a somewhat larger range (5.0–6.5 Å). The C $\beta$ —S—S—C $\beta$  and  $\chi^2$  torsion angles

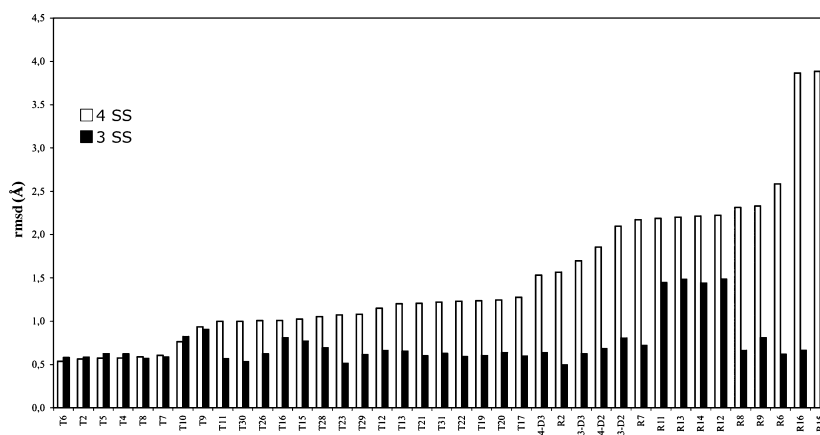
(N-terminal part of cystine C $\alpha$ —C $\beta$ —S—S) show that the majority of the former are confined to two regions (see supplementary Fig. S3), namely  $\pm 90^\circ$ , whereas the latter are contained within  $\pm 60^\circ$  to  $\pm 100^\circ$ , a region that is the typical range for such torsion angles [69]. There are several cases in which these angles deviate largely from the usual values, such as those derived from some of the NMR-established structures.

### On the structural conservation of the cystine cluster

The degrees of spatial variation of the three strictly conserved cystines B1, B2 and B4 that form the tight cluster described above and the less conserved B3 were calculated. To this end, we superimposed the three cystines from the  $\alpha$ -toxin of *N. nigricollis* (1IQ9), taken as a reference, on those of each of the other TFPDs established in crystallographic studies. As shown in Fig. 7 (black bars), the overall rmsd values vary from 0.5 to 1 Å, with a large majority having an rmsd close to 0.5 Å. For four TFPDs only, the rmsd value is close to 1.5 Å. This applies to the ECDs of some binary (1REW, 1ES7) and ternary (2H64, 2GOO) complexes of the receptors with the cytokines. We calculated the partial rmsd values for each atom in the B1, B2, B3 and B4 assembly, and found that, in the binary complex (1REW) and ternary complex (2H64), some large deviations are caused by the atoms in B1 and B3. It must be stressed that these structures are of bound receptors, and thus the diverse modes of binding between the cytokines and their ligands may account for the observed structural deviation [58]. In the other complexes, 3SS is highly affected (1S4Y, 1LX5 or 1KTZ). This was also observed, to a lesser extent, when free fasciculin (1FAS) was compared with its bound form (1FSS). We conclude that the overall spatial organization of the cystine cluster is highly



**Fig. 7.** The rmsd values calculated pairwise for the cystine network in 1IQ9, used as reference, and the cystine networks in the remaining TFPDs; black bars correspond to the three disulfide bridges B1, B2 and B4, and white bars correspond to the sets of four conserved disulfide bridges (B1, B2, B3 and B4). Data were sorted according to the increasing rmsd values in the 4S–S set of data. The abscissa indicates the indices given in Table 1.



conserved in TFPDs, and that this is unrelated to the functions of TFPDs, as the conservation is observed in TFPDs that act as ligands and receptors. This spatial organization of certain disulfides in the highly conserved SS network is affected, however, by binding of the ligands to the ECDs of the receptors.

We then examined the variations of spatial positioning of cystine B3 with respect to B1, B2 and B4. As shown in Fig. 7 (white bars), rmsd values below 0.7 Å were obtained for the short-chain neurotoxins that bind to postsynaptic acetylcholine receptors [43] and fasciculins that bind to acetylcholinesterase [25]. The rmsd values increased to about 1 Å for the toxins that bind to GPCRs, such as the long-chain neurotoxins that bind to both postsynaptic and neuronal acetylcholine receptors [43] in a species-specific manner [31], and for the cardiotoxins [32–40]. Variations in the rmsd values in the range 1.5–3.9 Å were observed exclusively for the TFPDs acting as ECDs. Therefore, there seems to be a trend which suggests that cystine B3, whose function is to lock the third finger (F3) in the TFPDs, is structurally less conserved, especially in the ECDs of receptors. Changing the spatial positioning of B3 with respect to the other three conserved disulfide bridges may illustrate some structural flexibility of TFPD, and could account for its adaptation to diverse biological functions.

### Diversified interaction modes between TFPDs and their ligands

For several binary and ternary complexes of the TFPDs listed in Table 1, we calculated all of the intermolecular contacts below 4.5 Å (see supplementary material). The networks of amino acids involved in the formation of diverse TFPD–ligand complexes are listed in supplementary Table S2. The complexes can be

divided into three groups: (1) T9, T10, T28 and T29, which consist of interactions between toxins and enzymes (T9 and T10) and mimics of the acetylcholine receptors (T28 and T29); (2) uPAR complexes (R3, R4 and R5) (whose TFPD-I and TFPD-II have the largest number of contacts, whereas TFPD-III has a small number of contacts) and their diverse ligands; and (3) the binary and ternary complexes between the TGFβ family of receptors and their different ligands. The interfaces in the binary and ternary complexes are mainly filled by side chain atoms with prevailing hydrophobic character. Some of these complexes are stabilized by hydrogen bonds [43,53,63] (see supplementary material).

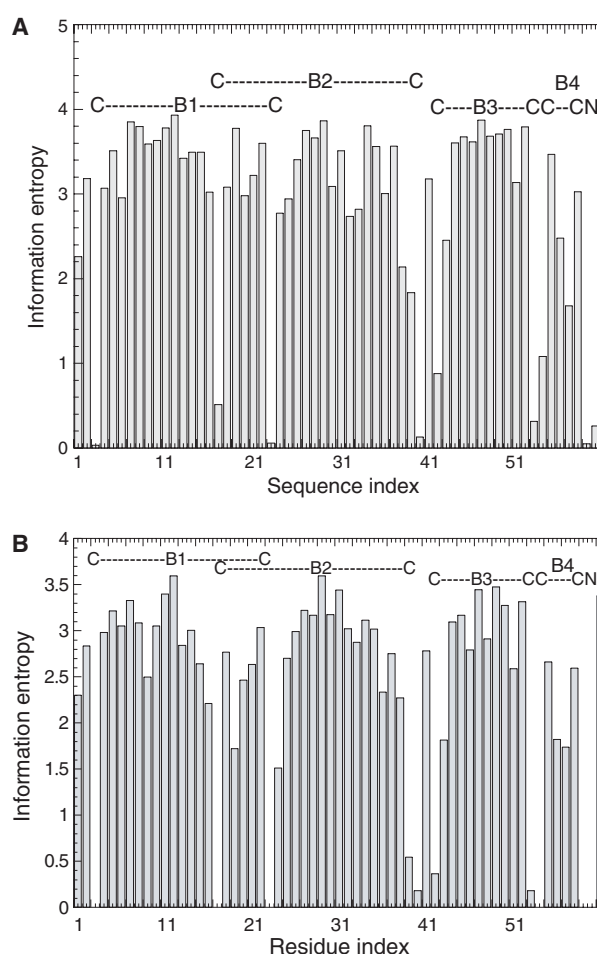
Toxins are bound to their receptors via the tips of F1, F2 and Lk3. It is worth noting, however, that the overall architecture of the fingers in diverse snake toxins displays less diversity of fine structural traits in comparison with the architecture of the fingers in the TFPDs of uPAR and the ECDs of the receptors. The last two groups have longer loops which are often flanked by short α-helices. Moreover, central finger 2 is longer in the ECDs of the receptors and the TFPDs of uPAR than it is in the snake toxins. For example, the ligands bind to uPAR in a deep cavity formed by the three consecutive TFPDs. Even in the complex of human uPAR with an antagonist peptide (1YWH), the three TFPDs form multiple contacts with the 13 amino acids of the antagonist. The group comprising the ECDs of the receptors displays an even wider range of interaction modes with the diverse ligands (see supplementary Table S2). Analyses of the ternary complexes Act-RIIB/BMP-RIA/BMP-2 (2H64) [62] and Act-RIIA/BMP-RIA/BMP-2 (2GOO) [63] revealed that the homodimeric BMP-2 ligand binds symmetrically two pairs of BMP-RIA and Act-RIIA (2GOO), and BMP-RIA and Act-RIIB (2H64). Although the

ECD of BMP-RIA does not interact with the ECD of Act-RIIA or Act-RIIB, in the structure 2GOO some interactions occur between the ECDs of the Act-RIIB units (see supplementary material). It has been concluded that the specific signalling output is dependent on at least two factors: (1) the specificity of the interactions between the homodimeric ligand BMP-2 and the ECDs; and (2) the way in which the dimeric receptor is assembled [63]. Such a scenario, however, would lead to a relatively large number of combinations of how diverse dimeric cytokines [55] may interact with the 12 different TGF $\beta$ -like receptors encoded in the human genome.

### Global analyses of sequences of diverse TFPDs

In order to examine whether or not the overall conclusions deduced from the analysis based on the selected set of protein sequences derived from the structures listed in Table 1 were valid, we analysed a larger set of sequences, including diverse toxins and ECDs of receptors extracted from several protein databases. The MSA was structured in the following fashion. At the top were grouped the sequences corresponding to the neurotoxins, cardiotoxins and weak neurotoxins from different snakes; the longer sequences corresponding to various ECDs and soluble forms of TFPDs were added to the bottom part of the MSA. Calculated distributions of the IDs revealed that the majority were in the range 10–20% for the 660 TFPDs (MSA660S1), whereas the peak moved to 20–30% for the set of snake toxins (data not shown). A simple way of showing the positional conservation of amino acids is the information entropy ( $I_e$ ) measure, as illustrated for the 660 TFPDs in Fig. 8A and the 36 unique sequences of the TFPDs assembled in Table 1 in Fig. 8B.

In general, the cysteine residues and the C-terminal asparagine are characterized by zero entropy (or close to) values, which confirms that these sequence positions are fully conserved (Fig. 8). In contrast, the other sequence positions can be highly variable, in particular in the finger regions. In MSA660S1 (see supplementary material), the cysteine residues involved in the formation of the cystine cluster (B1, B2 and B4) are characterized by an  $I_e$  value of 0.0, whereas the cysteine residues forming B3 are characterized by a slightly higher value (see supplementary material). Apart from these amino acids and the TFPD C-terminal CN doublet, overall sequence conservation is low amongst the diverse TFPDs. This is the result of several factors. Firstly, MSA660S1 includes several groups of TFPDs having different biological functions, which imply considerable sequence diversity. Secondly, gaps that



**Fig. 8.** Information entropy ( $I_e$ ) for the 660 TFPDs (A) and for the 36 unique sequences aligned in Fig. 2 (B).

were imposed by the different sequence lengths of the TFPDs perturbed the MSAs. Therefore, the 36 sequences used for structural alignment are equally diverse as the 660 sequences in MSA660S1. The general conclusion that commonly emerges from both analyses is that B1, B2, B4 and the C-terminal asparagine residues constitute virtually the strictly conserved structural cluster of the TFPDs, which could become a sufficient criterion for the database search for TFPD-like sequences. We suggest that the formation of the fine spatial organization of the cystine cluster may constitute a critical step during the folding process of TFPDs.

### Proteins with similar structural features to those in TFPDs

WGA and several plant lectins, such as hevein, have been shown previously to share similar structural

features to erabutoxin A, a typical TFPD, suggesting that these two types of protein adopt the 'snake toxin fold' [16]. The polypeptide chain of WGA comprises 171 amino acids and is composed of four consecutive units that have similar conformations. The sequence similarities between each WGA domain and erabutoxin B vary from 17% to 21%. However, a closer inspection of the amino acid sequences and three-dimensional structures shows that these lectins possess marked differences from TFPDs. As a consequence, WGA has been classified into a different knottin subgroup from the toxins [15].

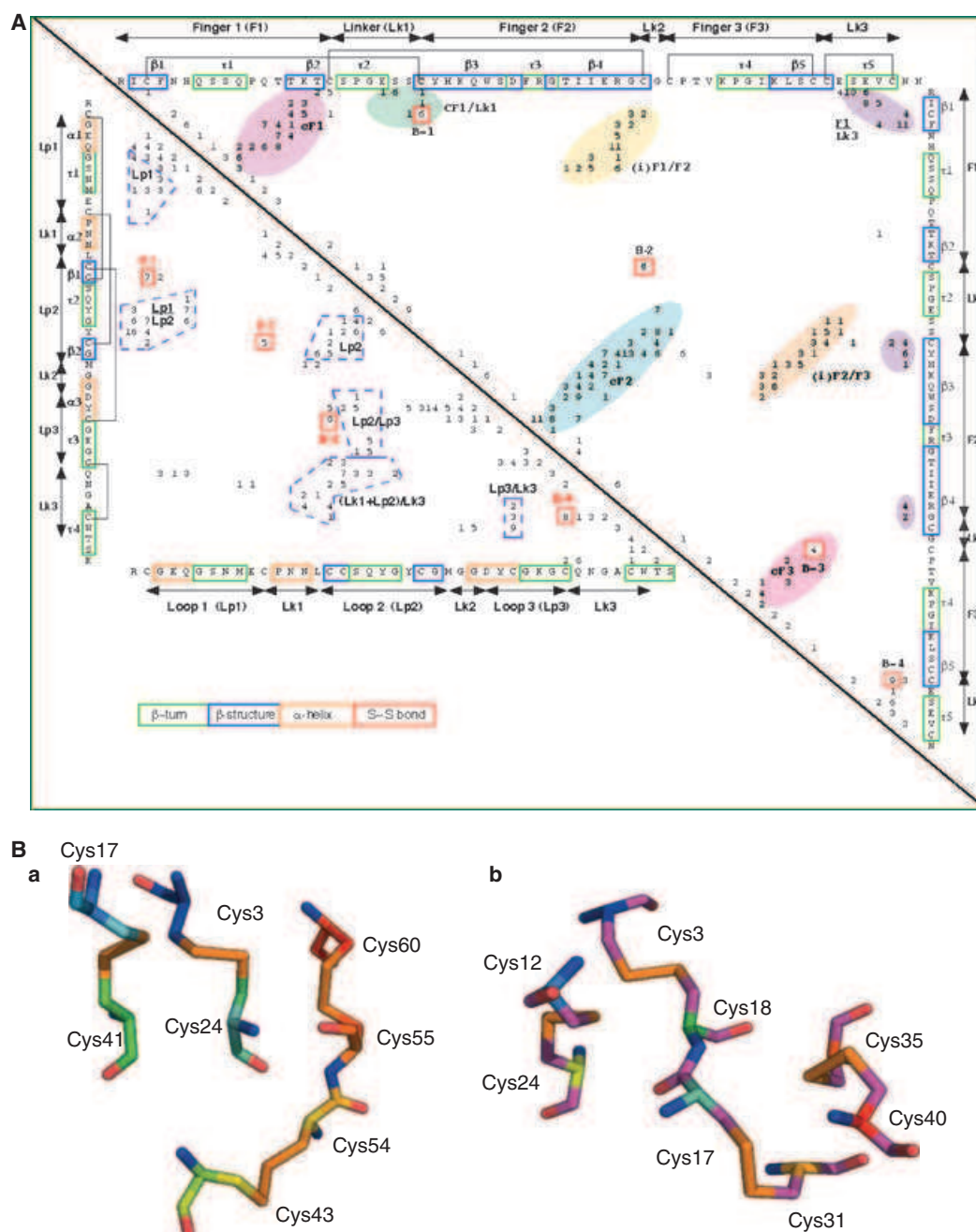
Firstly, WGA is characterized by loops composed of much shorter stretches (two amino acids) of  $\beta$ -strands,  $\beta$ -turns and short  $\alpha$ -helices (see Fig. 9A). Moreover, one domain of WGA is about 25% shorter than the smallest TFPD represented by dendroaspisin (59 amino acids) [16]. Secondly, WGA has no conserved asparagine residue adjacent to the second half-cystine of B4. Thirdly, the C-terminal loop (Lk3) shows a markedly different structural orientation in the TFPDs and some plant lectins. If we look at the structures from the side of the palm, Lk3 is oriented to the right in TFPDs and to the left in WGA. The lack of an asparagine after the second half-cystine could be at the origin of this marked deviation. Fourthly, a comparison of the distance maps of WGA-IV and a typical TFPD reveals that only the segments that are close to the first two cystines B1 and B2 have some resemblance in both WGA and TFPDs (see Fig. 9B). The B1, B2 and B4 bridges in TFPDs have an organization that can be compared with those seen in WGA. However, if the first two S-S bonds in the fourth repeats of WGA (9WGA) and erabutoxin B (6EBX) are superimposed, their rmsd value is 1.1 Å; superimposition of B1 and B3 gives rmsd = 3.8 Å, the three S-S bond combinations B1, B2 and B3 give rmsd = 3.4 Å, and B1, B2 and B4 give rmsd = 3.60 Å. Thus, only the spatial organization of B1 and B2 is shared between the TFPDs and the structurally similar plant lectins. Therefore, we conclude that proteins such as lectins do not belong to the TFPD family.

## Conclusions

This study has aimed to tentatively identify the structural determinants that are associated with the small protein domains called TFPDs which act as ligands, mainly toxins, or as the ECDs of some receptors. To this end, we analysed several hundred sequences containing TFPD-like motifs and 50 three-dimensional structures of diverse TFPDs. Firstly, the analysis

revealed that only the three disulfides B1, B2 and B4, and the asparagine that is adjacent to the second half-cystine of B4, are strictly conserved in the TFPDs. As many as 660 amino acid sequences from the genomes of diverse species were found to share the same conserved features, indicating that this fold has a wide distribution in the eukaryotic kingdom. Secondly, the conserved amino acid residue was found to be associated with the common presence of nine clusters of interactions and five  $\beta$ -strands organized into two  $\beta$ -pleated sheets composed of two or three strands. Interestingly, the largest number of contacts and the best energy terms were a result of these conserved half-cystines and a number of amino acids in their vicinity. In other words, the strictly conserved cystines B1, B2 and B4 and some adjacent amino acids are involved in large numbers of atomic contacts and provide important energy contributions. Therefore, we suggest that these amino acids are major stabilizing factors in the TFPD fold. Thirdly, a deeper analysis of the structure of the TFPDs revealed particularly strong interactions between B1/B2 and B1/B4 and between the conserved C-terminal asparagine region and B1 and B4. Therefore, we conclude that the assembly comprising B1/B2, B2/B4 and B2/B4/asparagine constitutes the principal stabilizing cluster of TFPDs. Several other components are highly, but not 100%, conserved in the TFPDs. This is the case in particular for the disulfide B3, which is lacking in several TFPDs. This disulfide also establishes a substantial number of interactions with neighbouring amino acids. Most interestingly, B3 shows substantial altered spatial positioning with respect to the conserved cystine clusters of TFPDs that act as ligands or receptors. The spatial orientation of B3 may therefore constitute a functional trait that differentiates the TFPDs from each other. Fourthly, high rmsd values were obtained when comparing the structures of other proteins that share distant structural resemblance with the TFPDs, namely the plant lectins WGA and hevein [16]. A deeper analysis of these two groups of proteins indicates that the lectins only share a common spatial organization with B1 and B2 of the TFPDs, strongly suggesting that these small proteins do not belong to the TFPD-like fold. The results presented in this article may be useful for future studies aiming to understand the folding mechanisms of diverse TFPDs, the phylogenesis of structurally related proteins [70,71], function-gain driven diversification of protein folds and the large functional diversity associated with the TFPD fold and other disulfide-rich small protein domains.





**Fig. 9.** (A) Bi-triangular distance maps for erabutoxin B (6EBX, top triangle) and the fourth domain in WGA (9WGA, bottom triangle). (B) Spatial arrangement of the disulfides in erabutoxin B (6EBX) and the fourth domain of WGA (9WGA).

## Experimental procedures

### Databases and sequence homology searching processes

The databases produced at the National Center of Biotechnology Information (NCBI) (<http://ncbi.nlm.nih.gov>) [72] and the Protein Information Resources (PIR) (<http://>

[pir.georgetown.edu](http://pir.georgetown.edu)) [73] were used in searches for diverse sequence motifs typical of the TFPDs.

### MSAs and their analyses

The DATA\_SQ program [74] was used to select diverse sets of sequences that were aligned with the CLUSTALW60 program [75] using the Blosom30 amino acid exchange matrix

[76] and a gap penalty set to 10. The quality of the MSAs was assured using the following rules: (1) the MSAs were manually adjusted according to the interaction patterns obtained from the analyses of three-dimensional structures, namely that the four canonical cystine bridges and the C-terminal CN doublet are well aligned; and (2) the sequence fragments that are between the canonical cystines were manually adjusted according to the physicochemical characteristics of the amino acids. The level of residue conservation at position  $j$  was estimated from the MSAs using the information (Shannon) entropy measure [77]:

$$I_j = - \sum_{a=1}^{20} p_{ja} \ln(p_{ja}) \quad (1)$$

where  $p_{ja}$  is the frequency of amino acids in column  $j$ .

### Structural analyses

The coordinates of X-ray structures were obtained from the Research Collaboration for Structural Bioinformatics (RCSB, <http://www.rcsb.org>) [1]. A suite of programs (CORDAN\_Prot) was derived from the original CORDAN program [78]. This suite was used to compute diverse geometry data and interatomic contacts from X-ray- and NMR-established structures of the TFPDs at a resolution of better than 3.3 Å, as described recently [79]. Briefly, distance maps were generated between all the atoms in the amino acids that were in  $i \geq i + 2$  sequence positions using two distance cut-offs, namely 4.0 and 4.5 Å. The calculated numbers of atomic distances were explicitly shown as integers on triangular maps that contained the amino acid sequences as coordinates. The amino acids with high Debye–Waller ( $B$  factor) values are shown in Table 1. The  $B$  factors were normalized using:

$$B(\text{norm}) = \{[B(j) - \text{Ave}]^2 / \sigma\} \quad (2)$$

where  $B(j)$  is the  $B$  factor of the  $j$ th amino acid residue, Ave is the average  $B$  factor and  $\sigma$  is its standard deviation. The numbers of interactions were normalized in the presentations of some graphs. Using eqn. (3), the bulkiness values of amino acids were divided by that of glycine [80], and this established scale was used for the normalization of the numbers of interactions per amino acid residue:

$$NI_i(\text{norm}) = NI_i / [(\text{bulkiness of glycine}) / (\text{bulkiness of amino acid } i)] \quad (3)$$

where  $NI_i$  is the number of distances of  $\leq 4.5$  Å between residue  $i$  and the other amino acids.

### Force field used and energy computing

Only the van der Waals' and Coulombic terms were employed in the calculation of the energy diagrams [79] using the AMBER protein force field [81]. The sum of the

van der Waals' and Coulombic energy terms for each amino acid residue was multiplied by 0.5, as the interaction energy is a result of different combinations of atoms in amino acids  $i$  and  $j$ . It has been shown that, as a result of a high density of packing of some interior atoms, their radii are somewhat shorter than those of the atoms in the exterior parts of proteins [82]. The  $\sigma$  factor of 0.9 was used to scale down the atomic radii. This was because the atomic distances between the sulfur atoms in the structures of the TFPDs were too short compared with the standard van der Waals' radius used in the AMBER force field. In Fig. S1, we show the average distance in the set of the conserved cystine motifs (B1, B2, B3 and B4), and the van der Waals' and Coulombic terms obtained for all the combinations of S–S distances of  $\leq 4.5$  Å were calculated.

### Cystine clusters

Clusters of sulfur atoms were established for all the analysed structures. The rmsd values were calculated by taking into account all 12 atoms of cystines and using the rotation/translation procedure developed by Kabsch [83]. We followed the propositions developed by Harrison and Steinberg [68] for computing the stacking (clusters) of cystines in the three-dimensional structure of proteins. The level of stacking (clustering) between two cystines was established in the following way: the distances between  $\alpha$ -carbon atoms in cystines A and B were calculated, namely  $CA^{\alpha 1}CB^{\alpha 1}$ ,  $CA^{\alpha 2}CB^{\alpha 1}$ ,  $CA^{\alpha 1}CB^{\alpha 2}$  and  $CA^{\alpha 2}CB^{\alpha 2}$ , where  $CA^{\alpha 1}$  is the  $\alpha$ -carbon of the N-terminal cysteine residue in cystine A,  $CB^{\alpha 1}$  is the same atom in the N-terminal cysteine residue in cystine B,  $CA^{\alpha 2}$  is the C-terminal cysteine residue in cystine A and  $CB^{\alpha 2}$  is the C-terminal cysteine residue in cystine B. A distance of 7.5 Å was used as cut-off. If three or four of these distances were in the range 3–7.5 Å, the clustering was considered to be high; if one or two of these distances were higher than 7.5 Å, the clustering was considered to be loose.

### References

- 1 Berman HM, Henrick K, Nakamura H & Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* **35**, D301–D303.
- 2 Andreeva A, Howorth D, Brenner SE, Hubbard TJP, Chothia C & Murzin AG (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res* **32**, D226–D229.
- 3 Levitt M & Gerstein M (1997) A structural census of the current population of protein sequences. *Proc Natl Acad Sci USA* **94**, 11911–11916.
- 4 Ouzounis CA, Coulson RM, Enright AJ, Kunin V & Pereira-Leal JB (2003) Classification schemes for protein structure and function. *Nat Rev Genet* **4**, 508–519.

- 5 Andreeva A & Murzin AG (2006) Evolution of protein fold in the presence of functional constraints. *Curr Opin Struct Biol* **16**, 399–408.
- 6 Grishin NV (2001) Fold change in evolution of protein structures. *J Struct Biol* **134**, 167–185.
- 7 Anantharaman V, Aravind L & Koonin EV (2003) Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr Opin Chem Biol* **7**, 12–20.
- 8 Arcus V (2002) OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr Opin Struct Biol* **12**, 794–801.
- 9 Larson SM & Davidson AR (2000) The identification of conserved interactions within the SH3 domain by alignment of sequences and structures. *Prot Sci* **9**, 2170–2180.
- 10 Low BW, Preston HS, Sato A, Rosen LS, Searl JE, Rudko AD & Richardson JS (1976) Three dimensional structure of erabutoxin b neurotoxic protein: inhibitor of acetylcholine receptor. *Proc Natl Acad Sci USA* **73**, 2991–2994.
- 11 Tsernoglou D & Petsko GA (1976) The crystal structure of a post-synaptic neurotoxin from sea snake at Å resolution. *FEBS Lett* **68**, 1–4.
- 12 Tsernoglou D & Petsko GA (1977) Three-dimensional structure of neurotoxin A from venom of the Philippines sea snake. *Proc Natl Acad Sci USA* **74**, 971–974.
- 13 Menez A (1998) Functional architectures of animal toxins: a clue to drug design? *Toxicon* **36**, 1557–1572.
- 14 Greenwald J, Fischer WH, Vale WW & Choe S (1999) Three-finger toxin fold for the extracellular ligand-binding domain of the type II activin receptor serine kinase. *Nat Struct Biol* **6**, 18–22.
- 15 Cheek S, Krishna SS & Grishin NV (2006) Structural classification of small, disulfide-rich protein domains. *J Mol Biol* **359**, 215–237.
- 16 Drenth J, Low BW, Richardson JS & Wright CS (1980) The toxin-agglutinin fold: a new group of small protein structures organized around a four-disulfide core. *J Biol Chem* **255**, 2652–2655.
- 17 Gilquin B, Bourgoïn M, Menez R, Le Du MH, Servent D, Zinn-Justin S & Menez A (2003) Motions and structural variability within toxins: implication for their use as scaffolds for protein engineering. *Prot Sci* **12**, 266–277.
- 18 Lou X, Liu Q, Tu X, Wang J, Teng M, Niu L, Schuller DJ, Huang Q & Hao Q (2004) The atomic resolution crystal structure of atratoxin determined by single wavelength anomalous diffraction phasing. *J Biol Chem* **279**, 39094–39104.
- 19 Cheng Y, Meng Q, Wang W & Wang J (2002) Structure–function relationship of three neurotoxins from the venom of *Naja kaouthia*: a comparison between the NMR-derived structure of NT2 with its homologues, NT1 and NT3. *Biochim Biophys Acta* **1594**, 353–363.
- 20 Gaucher JF, Menez R, Arnoux B, Pusset J & Ducruix A (2000) High-resolution x-ray analysis of two mutants of a curaremimetic snake toxin. *Eur J Biochem* **267**, 1323–1329.
- 21 Nastopoulos V, Kanellopoulos PN & Tsernoglou D (1998) Structure of dimeric and monomeric erabutoxin A refined at 1.5 Å resolution. *Acta Crystallogr D: Biol Crystallogr* **54**, 964–974.
- 22 Saludjian P, Prange T, Navaza J, Menez R, Guilloteau JP, Ries-Kautt M & Ducruix A (1992) Structure determination of a dimeric form of erabutoxin-B, crystallized from a thiocyanate solution. *Acta Crystallogr B* **48**, 520–531.
- 23 Le Du MH, Marchot P, Bougis PE & Fontecilla-Camps JC (1992) 1.9-Å resolution structure of fasciculin 1, an anti-acetylcholinesterase toxin from green mamba snake venom. *J Biol Chem* **267**, 22122–22130.
- 24 Le Du MH, Housset D, Marchot P, Bougis PE, Navaza J & Fontecilla-Camps JC (1996) Structure of fasciculin 2 from green mamba snake venom: evidence for unusual loop flexibility. *Acta Crystallogr D: Biol Crystallogr* **52**, 87–92.
- 25 Harel M, Kleywegt GJ, Ravelli RB, Silman I & Sussman JL (1995) Crystal structure of an acetylcholinesterase–fasciculin complex: interaction of a three-fingered toxin from snake venom with its target. *Structure* **3**, 1355–1366.
- 26 Kryger G, Harel M, Giles K, Toker L, Velan B, Lazar A, Kronman C, Barak D, Ariel N, Shafferman A et al. (2000) Structures of recombinant native and E202Q mutant human acetylcholinesterase complexed with the snake-venom toxin fasciculin-II. *Acta Crystallogr D: Biol Crystallogr* **56**, 1385–1394.
- 27 Menez R, Le Du MH, Gaucher JF & Menez A (2000) X-ray structure of muscarinic toxin 2 at 1.5 Å resolution. (<http://www.rcsb.org/pdb/>) [accessed in May 2005].
- 28 Kuhn P, Deacon AM, Comoso S, Rajaseger G, Kini RM, Uson I & Kolatkar PR (2000) The atomic resolution structure of buccandin, a novel toxin isolated from the Malayan krait, determined by direct methods. *Acta Crystallogr D: Biol Crystallogr* **56**, 1401–1407.
- 29 Murakami MT, Kini RM & Arni RK (2007) Crystal structure of buccain. (<http://www.rcsb.org/pdb/>) [accessed in October 2007].
- 30 Venkitakrishnan RP, Chary KVR, Kini MR & Govil G (2001) Solution structure of candoxin, a reversible, post-synaptic neurotoxin purified from the venom of *Bungarus candidus* (malayan krait). (<http://www.rcsb.org/pdb/>) [accessed October 2007].
- 31 Pawlak J, Mackessy SP, Fry BG, Bhatia M, Mourier G, Fuchart-Gaillard C, Servent D, Menez R, Stura EA, Menez A et al. (2006) Denmotoxin: a three-finger toxin from colubrid snake *Boiga dendrophila* (mangrove cat-snake) with bird-specific activity. *J Biol Chem* **281**, 29030–29041.
- 32 Bilwes A, Rees B, Moras D, Menez R & Menez A (1994) X-ray structure at 1.55 Å of toxin γ, a cardiotoxin from

- Naja nigricollis* venom. Crystal packing reveals a model for insertion into membranes. *J Mol Biol* **239**, 122–136.
- 33 Gilquin B, Roumestand C, Zinn-Justin S, Menez A & Toma F (1993) Refined three-dimensional solution structure of a snake cardiotoxin: analysis of the side-chain organization suggests the existence of a possible phospholipid binding site. *Biopolymers* **33**, 1659–1675.
- 34 Forouhar F, Huang WN, Liu JH, Chien KY, Wu WG & Hsiao CD (2003) Structural basis of membrane-induced cardiotoxin A3 oligomerization. *J Biol Chem* **278**, 21980–21988.
- 35 Wang C-H, Liu J-H, Lee S-C, Hsiao C-D & Wu W-G (2005) Glycosphingolipid-facilitated membrane insertion and internalization of cobra cardiotoxin: the sulfate/cardiotoxin complex structure in a membrane-like environment suggests a lipid-dependent cell-penetrating mechanism for membrane binding polypeptides. doi/10.1074/jbc.M507880200.
- 36 Chen TS, Chung FY, Tjong SC, Goh KS, Huang WN, Chien KY, Wu PL, Lin HC, Chen CJ & Wu WG (2005) Structural difference between group I and group II cobra cardiotoxins: X-ray, NMR, and CD analysis of the effect of cis-proline conformation on three-fingered toxins. *Biochemistry* **44**, 7414–7426.
- 37 Rees B, Bilwes A, Samama JP & Moras D (1990) Cardiotoxin VII4 from *Naja mossambica*: the refined crystal structure. *J Mol Biol* **214**, 281–297.
- 38 Sun YJ, Wu WG, Chiang CM, Hsin AY & Hsiao CD (1997) Crystal structure of cardiotoxin V from Taiwan cobra venom: pH-dependent conformational change and a novel membrane-binding motif identified in the three-finger loops of P-type cardiotoxin. *Biochemistry* **36**, 2403–2413.
- 39 Jayaraman G, Kumar TKS, Tsai CC, Chou SH, Ho CL & Yu C (2000) Elucidation of the solution structure of cardiotoxin analogue V from the Taiwan cobra (*Naja naja atra*) – identification of structural features important for the lethal action of snake venom cardiotoxins. *Prot Sci* **9**, 637–646.
- 40 Dementieva DV, Bocharov EV & Arseniev AS (1999) Two forms of cytotoxin II (cardiotoxin) from *Naja naja oxiana* in aqueous solution. Spatial structures with tightly bound water molecules. *Eur J Biochem* **263**, 152–162.
- 41 Betzel C, Lange G, Pal GP, Wilson KS, Maelicke A & Saenger W (1991) The refined crystal structure of  $\alpha$ -cobratoxin from *Naja naja siamensis* at 2.4-Å resolution. *J Biol Chem* **266**, 21530–21536.
- 42 Zeng H & Hawrot E (2002) NMR-based binding screen and structural analysis of the complex formed between  $\alpha$ -cobratoxin and an 18-mer cognate peptide derived from the  $\alpha$ 1 subunit of the nicotinic acetylcholine receptor from *Torpedo californica*. *J Biol Chem* **277**, 37439–37445.
- 43 Bourne Y, Talley TT, Hansen SB, Taylor P & Marchot P (2005) Crystal structure of a Cbtx-AChBP complex reveals essential interactions between snake  $\alpha$ -neurotoxins and nicotinic receptors. *EMBO J* **24**, 1512–1522.
- 44 Harel M, Kasher R, Nicolas A, Guss JM, Balass M, Fridkin M, Smit AB, Brejc K, Sixma TK, Katchalski-Katzir E *et al.* (2001) The binding site of acetylcholine receptor as visualized in the X-ray structure of a complex between  $\alpha$ -bungarotoxin and a mimotope peptide. *Neuron* **32**, 265–275.
- 45 Nickitenko AV, Michailov AM, Betzel C & Wilson KS (1993) Three-dimensional structure of neurotoxin-I from *Naja naja oxiana* venom at 1.9 Å resolution. *FEBS Lett* **320**, 111–117.
- 46 Dewan JC, Grant GA & Sacchettini JC (1994) Crystal structure of  $\kappa$ -bungarotoxin at 2.3 Å resolution. *Biochemistry* **33**, 13147–13154.
- 47 Moise L, Piserchio A, Basus VJ & Hawrot E (2002) Structural analysis of  $\alpha$ -bungarotoxin and its complex with the principal  $\alpha$ -neurotoxin-binding sequence on the  $\alpha$ 7 subunit of a neuronal nicotinic acetylcholine receptor. *J Biol Chem* **277**, 12406–12417.
- 48 Connolly PJ, Stern AS & Hoch JC (1996) Solution structure of ISIII, a long neurotoxin from the venom of *Laticauda semifasciata*. *Biochemistry* **35**, 418–426.
- 49 Sutcliffe MJ, Jaseja M, Hyde EI, Lu X & Williams JA (1994) Three-dimensional structure of the RGD-containing neurotoxin homologue, Dendroasp. *Nat Struct Biol* **1**, 802–807.
- 50 Fletcher CM, Harrison RA, Lachmann PJ & Neuhaus D (1994) Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Structure* **2**, 185–199.
- 51 Huang Y, Fedarovich A, Tomlinson S & Davies C (2007) Crystal structure of CD59: implications for molecular recognition of the complement proteins C8 and C9 in the membrane-attack complex. *Acta Crystallogr D* **63**, 714–721.
- 52 Llinas P, Le Du MH, Gardsvoll H, Dano K, Ploug M, Gilquin B, Stura EA & Menez A (2005) Crystal structure of the human urokinase plasminogen activator receptor bound to an antagonist peptide. *EMBO J* **24**, 1655–1663.
- 53 Huai Q, Mazar AP, Kuo A, Parry GC, Shaw DE, Callahan J, Li Y, Yuan C, Bian C, Chen L *et al.* (2006) Structure of human urokinase plasminogen activator in complex with its receptor. *Science* **311**, 656–659.
- 54 Barinka C, Parry G, Callahan J, Shaw DE, Kuo A, Bdeir K, Cines DB, Mazar A & Lubkowski J (2006) Structural basis of interaction between urokinase-type plasminogen activator and its receptor. *J Mol Biol* **363**, 482–495.
- 55 Allendorph GP, Iseacs MJ, Kawakami Y, Belmonte JC & Choe S (2007) BMP-3 and BMP-6 structures

- illuminate the nature of binding specificity with receptors. *Biochemistry* **46**, 12238–12247.
- 56 Greenwald J, Groppe J, Gray P, Wiater E, Kwiatkowski W, Vale W & Choe S (2003) The BMP7/ActRII extracellular domain complex provides new insights into the cooperative nature of receptor assembly. *Mol Cell* **11**, 605–617.
  - 57 Greenwald J, Vega ME, Allendorph GP, Fischer WH, Vale W & Choe S (2004) A flexible activin explains the membrane-dependent cooperative assembly of TGF- $\beta$  family receptors. *Mol Cell* **15**, 485–489.
  - 58 Thompson TB, Woodruff TK & Jardetzky TS (2003) Structures of an ActRIIB:activin A complex reveal a novel binding mode for TGF- $\beta$  ligand:receptor interactions. *EMBO J* **22**, 1555–1566.
  - 59 Mace PD, Cutfield JF & Cutfield SM (2006) High-resolution structures of the bone morphogenetic protein type II receptor in two crystal forms: implications for ligand binding. *Biochem Biophys Res Commun* **351**, 831–838.
  - 60 Keller S, Nickel J, Zhang JL, Sebald W & Mueller TD (2004) Molecular recognition of BMP-2 and BMP receptor IA. *Nat Struct Mol Biol* **11**, 481–488.
  - 61 Kirsch T, Sebald W & Dreyer MK (2000) Crystal structure of the BMP-2–BR1A ectodomain complex. *Nat Struct Biol* **7**, 492–496.
  - 62 Weber D, Kotsch A, Nickel J, Harth S, Seher A, Mueller U, Sebald W & Mueller TD (2007) A silent H-bond can be mutationally activated for high-affinity interaction of BMP-2 and activin type IIB receptor. *BMC Struct Biol* **7**, 6.
  - 63 Allendorph GP, Vale WW & Choe S (2006) Structure of the ternary signaling complex of a TGF- $\beta$  superfamily member. *Proc Natl Acad Sci USA* **103**, 7643–7648.
  - 64 Boesen CC, Radaev S, Motyka SA, Patamawenu A & Sun PD (2002) The 1.1 Å crystal structure of human TGF- $\beta$  type II receptor ligand binding domain. *Structure* **10**, 913–919.
  - 65 Hart PJ, Deep S, Taylor AB, Shu Z, Hinck CS & Hinck AP (2002) Crystal structure of the human T $\beta$ R2 ectodomain–TGF- $\beta$ 3 complex. *Nat Struct Biol* **9**, 203–208.
  - 66 DeLano WL (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA. (<http://pymol.sourceforge.net>).
  - 67 Brylinski M, Konieczny L & Roterman I (2006) Hydrophobic collapse in (*in silico*) protein folding. *Comp Biol Chem* **30**, 255–267.
  - 68 Harrison PM & Steinberg MJ (1996) The disulphide beta-cross: from cystine geometry and clustering to classification of small disulphide-rich protein folds. *J Mol Biol* **264**, 603–623.
  - 69 Srinivasan N, Sowdhamani R, Ramakrishnan C & Balaram P (1990) Conformations of disulfide bridges in proteins. *Int J Peptide Protein Res* **36**, 147–153.
  - 70 Ohno M, Menez R, Ogawa T, Danse JM, Shimohigashi Y, Fromen C, Ducancel F, Zinn-Justin S, Le Du MH, Boulain JC *et al.* (1998) Molecular evolution of snake toxins: is the functional diversity of snake toxins associated with a mechanism of accelerated evolution? *Prog Nucleic Acid Res Mol Biol* **59**, 307–364.
  - 71 Fry BG (2005) From genome to ‘venome’: molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res* **15**, 403–420.
  - 72 Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequiera E, Tatusova TA *et al.* (2003) Database resources of National Center for Biotechnology. *Nucleic Acids Res* **31**, 28–33.
  - 73 Wu CH, Yeh LSL, Huang H, Arminski L, Castro-Alvares K, Chen Y, Hu Z, Kourtesis P, Ledley RS, Suzek BE *et al.* (2003) The protein information resource (PIR). *Nucleic Acids Res* **31**, 345–347.
  - 74 Galat A (2004) Function-dependent clustering of orthologues and paralogues of cyclophilins. *Proteins* **56**, 808–820.
  - 75 Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
  - 76 Henikoff S & Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* **89**, 10915–10919.
  - 77 Arndt C (2004) *Information Measures: Information and its Description in Science and Engineering*. Springer, Berlin/Heidelberg.
  - 78 Galat A (1989) Analysis of dynamics trajectories of DNA and DNA–drug complexes. *CABIOS* **5**, 271–278.
  - 79 Galat A (2008) Functional drift of sequence attributes in the FK506-binding proteins (FKBPs). *J Chem Inf Mod* **48**, doi://10.1021/ci700429n.
  - 80 Tsai J, Taylor R, Chothia C & Gerstein M (1999) The packing density in proteins: standard radii and volumes. *J Mol Biol* **290**, 253–266.
  - 81 Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson M, Spellmeyer DC, Fox T, Caldwell JW & Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **117**, 5179–5197.
  - 82 Harpaz Y, Gerstein M & Chothia C (1994) Volume change on protein folding. *Structure* **2**, 641–649.
  - 83 Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr A* **32**, 922–923.

## Supplementary material

The following supplementary material is available online:

**Fig. S1.** Average distances in the disulfide network B1, B2, B3 and B4 (*y*-axis) vs. average (van der Waals' + Coulombic) energy terms calculated for the S–S networks of the structures shown in Table 1.

**Fig. S2.** Plot of the distributions of the distances between the pairs of  $C\alpha(r_{ij}^a)$  and  $C\beta(r_{ij}^b)$  atoms of cystines in the chosen set of TFPDs.

**Fig. S3.** Plot of the distribution of the  $C\beta$ –S–S– $C\beta$  torsion angle (*x*-axis) vs. the  $C\alpha$ – $C\beta$ –S–S torsion angle (*y*-axis).

**Table S1.** Numbers of interactions in some sequence motifs in the TFPDs.

MSA of 660 TFPDs and associated sequence attributes file (MSA660.S1, MSA.S1.out).

**Table S2.** Intermolecular distances in several binary and ternary complexes involving different TFPDs (see Table 1 and Interfaces.S3.out file).

TFPD660.S4.out and TFPDXray.S5.out contain numerical values of  $I_e$ .

This material is available as part of the online article from <http://www.blackwell-synergy.com>

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

---

## **ANNEXE 3**

---

Séquences des oligonucléotides





Oligos mutagenèse ponctuelle:

Toxa 62

GGTTTCACCAGGGGCTACAGGTTTTAGT

C43A

CGTGATGTGGCGCTCCTACCGTTAAACCT

C54A

CCTGGTATCAAGCTTAACGCTTGCACCACTGAT

Oligo d'amplification

pCPM62

GTGACATGAGCTAGATGCAATGGTACCGatgctggaatgccacaacc

Oligos muta polyS

Oligo1 (CysSer1)

GTGACATGAGCTAGATGCAATGGTACCGATGCTGGAATCCCCACAACC

Oligo2 (CysSer2)

GGAGTGGTTTTACCAGGGCTAGAGGTTTTAGT

Oligo 3 (CysSer3)

ACCTCTAGCCCTGGTGAAACCAACTTCCTTACAAAAAGGTT

Oligo 4 (CysSer4)

AACGGTAGGGGGAGCCAGATTCCACGCTC

Oligo 5 (CysSer4bis)

GAGCGTGGATCTGGCTCCCCTACCGTT

Oligo 6 (CysSer5)

AGTACCTTATGCTTGGACTGGGATCCTTAATTGTTAGATTTATCAGTGGT  
GGAAGAGTTAAGC

Oligo insertion ETagEtag

TxEtag

gtctgggtacggaaccggagcaccATTGTTACATTTATCAGTGGTG

EtagBamH1

TTAGTTAGTACCTTATGCTTGGACTGGGATCCTTAgtctgggtacggaac  
cgg

SETag

gtctgggtacggaaccggagcaccATTGTTAgATTTATCAGTGGTG

---

## **ANNEXE 4**

---

Comparaison des contacts N.O.E. observés :

- entre Tox61 et Tox61pT3
- entre Tox62 et Tox62pT3



Asn22		
Atome ou groupement	Tox61	Tox61pT3
hn	Leu1 hg, hd1, hd2 Glu20 hg Thr21 hn Cys23 hb Gly41 ha	Superpositions
ha	Leu1 hd1, hd2 Thr21 hg Cys23 hn Thr55 ha	Leu1 hd1, hd2  Cys23 hn Thr55 ha
hb	Cys23 hn Cys42 hn  Cys54 hn	Cys23 hn Ala42 hb Ala53 hb
hnd21, hnd22	Thr21 hb, hg Cys42 hb1, hb2 Cys53 hb2	Thr21 hg Ala42 hb Ala53 hb

Cys42 ou Ala42		
Atome ou groupement	Tox61	Tox61pT3
hn	Asn22 hb1 Pro43 ha, hd1 Gly41 hn, ha1, ha2	Superpositions
ha	Tyr24 hb2 Pro43 hd1, hd2 Gly41 hn	Eau
hb	Asn22 hnd21, hnd22	Asn22 hb1, hb2 hnd21, hnd22 Pro 43 hd1

Cys53 ou Ala53		
Atome ou groupement	Tox61	Tox61pT3
hn	Asn52 ha,hb1, hb2 Cys54 hn	Asn52 ha, hb1, hb2
ha	Tyr24 ha, hb1 Lys25 hn Cys54 hn	Tyr24 ha  Cys54 hn
hb	Asn22 hb1, hb2  Cys54 hn	Asn22 hb2, hnd21, hnd21 Tyr24 ha Cys54 hn

Thr21		
Atome ou groupement	Tox61	Tox61pT3
hn	Leu1 hd1 Glu20 hb1, hb2, hg Asn22 hn	Leu1 hd1 Glu20 ha, hb1, hb2, hg Asn22 hn
ha, hb	Leu1 hd1, hd2	Leu1 hd1, hd2
hg		

Tyr24		
Atome ou groupement	Tox61	Tox61pT3
hn	Cys23 ha, hb1, hb2  Gly41 hn  Pro43 hd Asn60 hnd21, hnd22	Cys23 ha, hb1, hb2 Gly39 hn  Ala42 ha  Asn60 hnd21, hnd22
ha	Cys23 hb1, hb2 Lys25 ha Lys26 hn Arg38 hb  Asn60 hnd21, hnd22	Lys25 hn   Ala42 ha Ala53 ha, hb Cys54 hn
hb	Lys25 hn Cys42 ha Cys53 ha	Lys25 hn  Ala53 hb

Val27		
Atome ou groupement	Tox61	Tox61pT3
hn	Lys26 ha  Ile49 hg1, hg21 Lys50 hn, hb1	Lys26 ha Ile36 ha Ile49 hg1, Lys50 hb1
ha	Ile36 hg1, hg21	Trp28 ha, he3
hb	Ambigu	
hg	Lys25 he2  Trp28 ha Ile36 ha	Lys26 ha Trp28 ha Ile36 ha

Thr44		
Atome ou groupement	Tox61	Tox61pT3
hn	Pro43 ha, hb1, hb2, hg1, hg2 Val45 hn	Pro43, hb1, hb2 Val45 hn, ha
ha	Val45 hn, hg1, hg2	Val45 hn
hb	Val45 hn	Val45 hn
hg	Val45 hn	Val45 hn

Val45		
Atome ou groupement	Tox61	Tox61pT3
hn	Thr44 ha, hb, hg	Ambigu (deux pics de la forme en échange)
ha	Lys46 hn	Ambigu
hb	Lys46 hn	Lys46 hn
hg	Lys46 hn	Lys46 hn

Lys50		
Atome ou groupement	Tox61	Tox61pT3
hn	Lys26 ha Trp28 hb1, hb2 Ile49 ha, hb, hg1	Trp28 hb2 Ile49 ha, hg1, hg21
ha	Ambigu	Ambigu
hb	Leu51 hn	Leu51, hn
hg, hd et he		

Leu 51		
Atome ou groupement	Tox61	Tox61pT3
hn	Ambigu avec gly48 hn et cys17 hn	Ambigu avec tyr24
ha	Lys26 ha Val45 hg2 Asn52 hn	Lys26 ha Val45 ha Asn52 hn
hb	Val45 hg1 Asn52 hn	Asn52 hn
hg	Tyr24 hd Asn52 ha	
hd	Tyr24 ha, hd Lys26 ha Cys42 ha Thr44 ha Val45 hn Asn52 ha	

Cys54		
-------	--	--

Atome ou groupement	Tox61	Tox61pT3
hn	Asn22 hb2 Cys23 hn Tyr24 ha Cys53 ha, hb1, hb2	Asn22 ha, hb2  Tyr24 ha Ala53 ha, hb
ha	Cys23 hn Thr55 hn, hg	Thr55 hn, hg
hb	Thr55 hn	Thr55 hn

Thr55		
Atome ou groupement	Tox61	Tox61pT3
hn	Cys54 ha, hb1, hb2 Thr56 hn, ha	Cys54 ha, hb1, hb2 Thr56 hn
ha	Asn22 ha	Asn22 ha
hb	Eau	Eau
hg	Ambigu	Ambigu

Glu20		
Atome ou groupement	Tox61	Tox61pT3
hn	Gly19 hn, ha1, ha2	Gly19 hn, ha1, ha2
ha	Eau	Eau
hb	Cys17 hb1, hb2	Cys17 hb1, hb2
hg	Leu1 hd Cys40 hb2	Leu1 hd Cys40 hb1, hb2

Asn 23		
Atome ou groupement	Tox62	Tox62pT3
hn	Nombreuses superpositions	Nombreuses superpositions
ha	Leu1 hd Thr56 ha et hg	Leu1 hd Thr56 ha et hg
hb	Cys24 hn Cys43 hb Cys54 hb	Cys24 ha Ala43 hb Ala54 hb Thr56 hg
hnd11 et hnd12	Thr22 hb Cys43 hb Cys54 hb Thr56 hg	Ala43 hb Ala54 hb Thr56 hg

Cys43 ou Ala43		
Atome ou groupement	Tox62	Tox62pT3
hn	Asn23 hb (super et ambigu) Gly42 hn, ha1 et ha2 Pro44 hd	Asn23 hb (ambigu) Gly42 ha1 et ha2 Pro44 hd
ha	Tyr25 hb Pro44 hd Leu52 hd	Asn23 hb (ambigu) Tyr25 hb et hd Pro44 hd Leu52 hd
hb	Asn23 hnd21, hnd22 Tyr25 hb2	Asn23 hb1, hb2 hnd21, hnd22 Tyr25 hb1 Pro44 ha, hd Ala54 hb

Cys54 ou Ala54		
Atome ou groupement	Tox62	Tox62pT3
hn	Nombreuses superpositions	Nombreuses superpositions
ha	Tyr25 ha, hb1 et hb2 Lys26 hn Cys43 hb1 Cys55hn	Tyr25 ha Lys26 hn
hb	Asn23 hb (Ambigu) Tyr25 ha, hb et hd  Cys55 hn	Asn23 hb Tyr25 ha Ala43 hb Cys55 hn



Thr56		
Atome ou groupement	Tox62	Tox62pT3
hn	Cys55 ha, hb Thr57 hn, ha	Cys55 ha, hb Thr57 hn
ha	Leu1 hd1 et hd2 Asn23 ha, hb1 et hb2 Thr57 hn	Leu1 hd1 Asn23 ha Thr57 hn
hb	eau	eau
hg	Asn23 ha hnd11 et hnd12 Cys55 ha Thr57 hn	Asn23 ha hnd11 et hnd12 Cys55 ha

Val46		
Atome ou groupement	Tox62	Tox62pT3
hn	Thr45 hn, ha, hb et hg Lys47 hn	Thr45 hn, ha, hb et hg Lys47 hn
ha	Thr45 hg Lys47 hn	Thr45 hn, ha Lys47 hn Leu52 hb, hg
hb	Lys47 hn	Lys47 hn Ile50 hg
hg	Thr45 ha Lys47 hn Ile50 ha Lys51 ha, hb	Thr45 ha Lys47 hn Ile50 hn, ha, hb, hd Lys51 ha, hb

Glu21		
Atome ou groupement	Tox62	Tox62pT3
hn	Superposition	Superposition
ha	Thr22 hn	Thr22 hn
hb		
hg	Asn23 hn Gly42 hn, ha1 Cys43 hn	Ambiguïté avec Asn23 hb

Tyr25		
Atome ou groupement	Tox62	Tox62pT3
hn	Cys24 ha, hb2 Arg39 ha Cys41 ha Asn61 hb, hnd21 et hnd22	Superposition avec Ala54 hn
ha	Lys26 hn <b>Arg39 ha</b> Leu 52 hd Cys54 ha Cys55 hn	Lys 26 hn  Leu 52 hd Ala54 ha, hb Cys55 hn <b>Asn61 hb</b>
hb	Lys26 hn <b>Cys43 ha</b> Leu 52 hd	Lys 26 hn  Leu 52 hd

Ile50		
Atome ou groupement	Tox62	Tox62pT3
hn	Val46 hg Gly49 hn, ha1 et ha2	Val46 hg Gly49 hn, ha1 et ha2
ha	Eau	Eau
hb	Ambigu	Ambigu
hg	<b>Trp29 hn, ha, hb1, hb2, he3</b>	<b>Val46 ha, hg</b>
hd	Trp29 ha, hb1, hb2, hd1, he3  Val 46 ha, hb Lys47 hn	Trp29 hb1, hb2 (bien plus faible intensité que pour Tox62) Val46 ha, hb, hg Lys 47 hn, hb

Leu52		
Atome ou groupement	Tox62	Tox62pT3
hn	Superposition avec gly49	Superposition avec Glu2
ha	Ambigu avec Cys41 ha	Lys 27 ha Asn53 hn
hb		
hg, hd	Tyr25 hd Lys27 ha Asn53 hn	Tyr25 ha, hb, hd Lys27 hn, ha Asn53 hn, hb

Thr45		
Atome ou groupement	Tox62	Tox62pT3
hn	Superposition avec gln6ne2 et Glu21	Pro44 ha, hb Val46 hn
ha	Val46 hn, hg	Pro44 ha Val46 hn, hg
hb	Ambigu	Ambigu
hg	Val46 hn	Val46 hn

Lys47		
Atome ou groupement	Tox62	Tox62pT3
hn	Val46 hn, hb, hg Ile50 hb, hg, hd Leu52 hg	Val46 hn, hb, hg Pro48 hg Ile50 hb, hd
ha	Eau	Eau
hb	Ambigu	Ambigu
hg	Ambigu	Ambigu

Lys51		
Atome ou groupement	Tox62	Tox62pT3
hn	Val28 hn, ha Ile50 hn, ha, hg	Val28 hn Ile50 ha hg
ha	Eau	Eau
hb	Val28 hn	
hg		

Gly42		
Atome ou groupement	Tox62	Tox62pT3
hn	Leu1 hd Glu21 hg Asn23 hn, hb1 hb2 Cys24 ha Tyr25 hd Cys41 ha, hb1, hb2 Cys43 hn	Glu21 hg  Cys24 ha Tyr25 hb, hd Cys41 ha, hb1, hb2
ha	Cys43 hn	Ala43 hn

Val28		
Atome ou groupement	Tox62	Tox62pT3
hn	Lys27 ha, hg, hd <b>Trp29 he3</b> Ile50 hg1, hg21 Lys51 hn, hb Leu52 ha	Lys27 ha Ile50 hg11, hg2 Lys51 hn, hb Leu 52 ha
ha	<b>Trp29 he3</b>	<b>Thr35 hg</b> <b>Ile37 hg11, hd</b> <b>Lys51 hn</b>
hb	Ambigu	Ambigu
hg	Ile37 ha	Ile37 ha

Thr22		
Atome ou groupement	Tox62	Tox62pT3
hn	Leu1 hd1 Thr56 hg2 Glu 21 ha, hb1 hb2 <b>Asn23 hn, ha, hb1</b>	Leu1 hd1 Asn23 hn, hb1 Glu21 ha
ha	Leu1 hd Asn23 hn	Leu1 hd Asn23 hn
hb	Ambigu	Ambigu
hg	Ambigu	Ambigu

Cys55		
Atome ou groupement	Tox62	Tox62pT3
hn	Asn23 hn, hb2 Cys 24 hn, hb1 Tyr25 ha Cys54 hn, hb1, hb2 Thr56 ha, hg2	Asn23 hb2 Cys 24 hn(superposé), hb1 Tyr25 ha Ala54 ha, hb Thr56 hn, hg
ha	Thr56 hn, hg2 Thr57 hn	<b>Ala54 hb</b> Thr56 hn, hg2 Thr57 hn
hb	Thr56 hn, hg2	Thr56 hn, hg2

Cys24		
Atome ou groupement	Tox62	Tox62pT3
hn	Leu1 hd Asn23 hb1, hb2 Cys54 ha Thr56 hg2	Leu1 hd Asn23 ha, hb1, hb2 Ala54 ha (superposé) Thr56 ha
ha	Tyr25 hn Cys41 ha Gly42 hn	Tyr25 hn Cys41 ha Gly42 hn
hb	Tyr25 hn Asn61 hb1 hnd21, hnd22	Tyr25 hn Asn61 hb1 hnd21, hnd22

---

## **ANNEXE 5**

---

Liste des déplacements chimiques de Tox61, Tox62, Tox61pT3 et Tox62pT3



## Liste des déplacements chimiques de Tox61

Tox61			n	Hn	Ha1	Ha2	Hb1	Hb2	Hg1	Hg2	Hd1	Hd2	He1	He2	Hcycle
Leu	L	1			4,14		1,54	-	1,45		0,61	0,69			
Glu	E	2	127,7	8,96	5		1,78	1,84	2,03						
Cys	C	3	121,1	8,73	5,13		2,34	2,91							
His	H	4	120,7	9,64	5,17		2,72	3,54							
Asn	N	5	116,5	8,38	4,86		2,1	3,06							nd:113,4 hnd1:6,99 hnd2:7,53
Gln	Q	6	126,5	9,12	4,55		1,93	-	2,64						ne:109,2 hne1:6,51 hne2:7,77
Gln	Q	7	122,2	8,45	4,82		2,02	2,16	2,57						ne:112,1 hne1:6,90 hne2:7,58
Ser	S	8	116,3	9,69	3,91		3,99	4,14							
Ser	S	9	119,7	8,31	4,73		3,97	4,03							
Gln	Q	10	121,2	8,06	4,36		2,03	2,31	2,48	2,64					ne:111,3 hne1:6,91 hne2:7,40
Pro	P	11			4,59		1,79	2,37	2,16		3,63	3,82			
Pro	P	12			-		-	-	-						
Thr	T	13	114	8,00	4,81		-	-	-						
Thr	T	14	112,8	8,58	5,08		3,8	-	0,5						
Lys	K	15	118,5	8,81	4,67		1,61	1,77	1,29	-	1,69	-	3,02	-	
Thr	T	16	117,8	8,49	4,53		4,04	-	1,16						
Cys	C	17	124,3	8,64	5,13		2,9	3,4							
Pro	P	18			4,36		1,96	2,29	2,12	-	-	-			
Gly	G	19	111,5	8,84	3,82	4,06									
Glu	E	20	118,9	7,33	4,7		1,91	2,3	2,4						
Thr	T	21	111,4	8,79	4,42		4,36		1,19						
Asn	N	22	119,5	8,26	5,64		2,31	2,6							ne:111 hnd1:6,72 hnd2:7,04
Cys	C	23	116,2	9,13	5,54		2,7	3,35							
Tyr	Y	24	113,7	8,66	6,23		2,95	3,84							hd:6,72 he:6,57
Lys	K	25	124,9	8,98	5,12		1,62	1,89	1,21		1,13	1,54	2,69	2,74	
Lys	K	26	130,7	9,98	5,51		1,91	2,09	1,52	-	1,38	-	2,46		
Val	V	27	123,9	9,15	5,33		2,03		0,92	0,99					
Trp	W	28	128,6	9,17	5,11		3,44	3,5							ne:130 hne1:10,27 hd1:6,97 hh2:7,18 hz3:6,87 hz2:7,41 he3:7,32
Arg	R	29	118,9	8,37	5,06		1,67	1,8	1,56		3,12				ne:117 hne:7,08
Asp	D	30	124,9	8,47	4,86		2,8	3,19							
His	H	31			-		-	-	-	-	-				
Arg	R	32	118,6	8,60	4,21		1,81		1,17	1,53	3,09	3,18			ne:116,7 hne7,43
Gly	G	33	108,3	7,89	3,78	4,31									
Thr	T	34	117,6	8,30	4,44		3,87		1,02						
Ile	I	35	131,4	9,02	3,89		0,12		0,66	0,84	0,34				gCH3:0,25
Ile	I	36	121,4	7,37	5,25		1,46		1,08	1,58		1,17			gCH3:0,81
Glu	E	37	129	9,51	4,96		2,28	2,32	2,62						
Arg	R	38	125,2	8,73	4,88		1,14		1,46	1,58	3,11				ne:117,1 hne:7,37
Gly	G	39	104,6	6,22	3,79	3,9									
Cys	C	40	118,7	8,75	5,16		3,15	3,22							
Gly	G	41	111,3	9,04	3,76	4,2									
Cys	C	42	113,5	8,13	5,14		2,56	2,89							
Pro	P	43			4,39		1,67	1,92	0,77	0,93	2,83	2,92			
Thr	T	44	111,7	7,75	4,17		3,97		1,18						
Val	V	45	118,9	8,27	4,43		2,07		0,76	0,84					
Lys	K	46	122,9	7,67	4,51		1,53	1,81	1,39	1,44	1,59	1,62	2,9		
Pro	P	47			4,35		1,84	2,34	2,01	2,11	3,56	3,82			
Gly	G	48	111,2	8,64	3,68	4,33									
Ile	I	49	121,3	7,51	4,38		1,9	-	0,92	1,26	0,43				gCH3:0,76
Lys	K	50	128	8,57	4,52		1,91	2	1,7	-	1,45	1,5	3,02		
Leu	L	51	129	8,64	5,28		1,24	1,86	1,46	-	0,85				
Asn	N	52	124,6	8,94	5,13		2,65	2,73							nd:111 hnd1:6,82 hnd2:7,49
Cys	C	53	122,5	9,05	5,85		3,18	3,51							
Cys	C	54	120,3	9,36	5,36		3,31	3,68							
Thr	T	55	108,3	8,42	5,06		4,81		1,21						
Thr	T	56	112,1	7,44	4,8		4,25		1,19						
Asp	D	57	119,3	8,27	-		-		-						
Lys	K	58	114	9,54	3,29		1,88	2,12	1,01	-	0,51	-	2,77		
Cys	C	59	113,2	7,82	4,5		3,38	3,83							
Asn	N	60	121,8	9,14	4,75		2,18	2,82							nd:117,6 hnd1:7,98 hnd2:8,25
Asn	N	61	121,3	7,51	3,9		2,38	2,57							nd:111,5 hnd1:6,39 hnd2:7,29



## Liste des déplacements chimiques de Tox62

Tox62			n	Hn	Ha1	Ha2	Hb1	Hb2	Hg1	Hg2	Hd1	Hd2	He1	He2	Hcycle
Leu	L	1			4,12		1,41		1,32		0,59	0,71			
Glu	E	2	128,3	8,72	5,15		1,74		1,94	2,05					
Cys	C	3	119,9	8,59	5,03		2,35	2,95							
His	H	4	120,5	9,57	5,17		2,72	3,54							
Asn	N	5	116,9	8,39	4,87		2,12	3,03							nd:113,3 hnd1:6,99 hnd2:7,52
Gln	Q	6	126,6	9,13	4,55		1,91		2,61						ne:109,2 hne1:6,51 hne2:7,77
Gln	Q	7	122,3	8,45	4,82		2,01	2,17	2,56						ne:112,1 hne1:6,90 hne2:7,57
Ser	S	8	116,3	9,69	3,92		3,99	4,13							
Ser	S	9	119,5	8,31	4,7		4	4,03							
Gln	Q	10	121,2	8,05	4,36		2,02	2,3	2,48	2,63					ne:111,3 hne1:6,91 hne2:7,40
Pro	P	11			4,58		1,79	2,36	2,16	-	3,63	3,82			
Pro	P	12			-		-	-	-	-	-	-			
Thr	T	13	114	8,03	4,8		-	-	-	-	-	-			
Thr	T	14	113,6	8,79	5,03		3,78		0,55						
Lys	K	15	119,1	8,77	4,64		1,57	1,73	1,23		1,67		2,98		
Thr	T	16	120,4	8,54	4,51		3,98		1,3						
Cys	C	17	125,9	8,73	4,83		2,77	3,58							
Ser	S	18	119,4	9,15	4,58		3,87	4,01							
Pro	P	19			4,33		1,9	2,33	2,05	2,15	3,69	3,96			
Gly	G	20	112,1	8,69	3,69	4,22									
Glu	E	21	121,6	7,76	4,28		2,03	2,22	2,5						
Thr	T	22	115,4	8,25	4,44		4,36		1,02						
Asn	N	23	121,2	9,04	5,69		2,31	2,6							nd:111,6 hnd1:6,69 hnd2:7,19
Cys	C	24	115,5	9,18	5,69		2,79	3,3							hd:6,72 he:6,55
Tyr	Y	25	113,6	8,79	6,21		2,94	3,87							
Lys	K	26	124,8	8,97	5,1		1,9		1,11	1,2	1,55	1,6	2,69	2,76	
Lys	K	27	130,8	9,95	5,51		1,89	2,1	1,38		1,48		2,42	2,49	
Val	V	28	124,1	9,18	5,33		2,03		0,92	0,98					hd1:6,98 ne1:130,0 he1:10,27 he3:7,32 hz2:7,40 hz3:6,87 hh2:7,19
Trp	W	29	128,6	9,19	5,11		3,44	3,49							ne:117 hne:7,08
Arg	R	30	118,8	8,37	5,06		1,67	1,81	1,55						
Asp	D	31	124,9	8,47	4,85		2,8	3,18			3,13				
His	H	32	-												
Arg	R	33	118,6	8,60	4,2		1,82		1,17	1,53	3,09	3,18			ne:116,7 hne:7,43
Gly	G	34	108,3	7,89	3,78	4,3									
Thr	T	35	117,6	8,30	4,44		3,87		1,02						
Ile	I	36	131,4	9,02	3,89		0,1		0,67	0,84	0,33				gCH3:0,25
Ile	I	37	121,4	7,38	5,26		1,46		1,07		0,81				gCH3:0,81
Glu	E	38	128,9	9,51	4,96		2,32	2,62							
Arg	R	39	125	8,75	4,86		1,14	1,23	1,47	1,57	3,1				ne:116,8 hne:7,30
Gly	G	40	104,7	6,23	3,81	3,93									
Cys	C	41	119,2	8,58	5,32		3,09	3,22							
Gly	G	42	112	9,33	3,8	4,3									
Cys	C	43	113	8,07	5,07		2,6	2,88							
Pro	P	44			4,35		1,91	2,31	0,83	0,98	2,75	2,88			
Thr	T	45	111,9	7,76	4,16		3,97		1,18						
Val	V	46	118,9	8,28	4,44		2,07		0,76	0,84					
Lys	K	47	122,8	7,65	4,5		1,53	1,81	1,4	1,62					
Pro	P	48			-		-	-	-						
Gly	G	49	111,3	8,65	3,68	4,33									
Ile	I	50	121,4	7,54	4,38		1,92		0,92	1,27	0,43				gCH3:0,77
Lys	K	51	128	8,59	4,52		1,9	2	1,7		1,46		3,08		
Leu	L	52	128,9	8,64	5,31		1,25	1,86	1,46		0,86				
Asn	N	53	124,5	8,91	5,13		2,64	2,74							nd:111,1 hnd1:6,83 hnd2:7,52
Cys	C	54	122,3	9,05	5,9		3,17	3,55							
Cys	C	55	120,3	9,43	5,34		3,34	3,7							
Thr	T	56	108,3	8,42	5,22		4,8		1,2						
Thr	T	57	112,4	7,48	4,77		4,28		1,2						
Asp	D	58	119	8,20	-		2,4	2,51							
Lys	K	59	114,1	9,50	3,29		1,86		0,48		0,98		-		
Cys	C	60	113,2	7,84	4,5		3,39	3,82							
Asn	N	61	121,9	9,14	4,76		2,2	2,83							nd:117,5 hnd1:7,98 hnd2:8,24
Asn	N	62	121,4	7,56	3,91		2,41	2,57							nd:117,7 hnd1:6,42 hnd2:7,32

## Liste des déplacements chimiques de Tox61pT3 forme A

Tox61pT3A		n	Hn	Ha1	Ha2	Hb1	Hb2	Hg1	Hg2	Hd1	Hd2	He1	He2	Hcycle
Leu	L	1		4.13		1.54	-	1.46	-	0.62	0.70			
Glu	E	2	127.6	8.97	5.02	1.79	1.83	2.03						
Cys	C	3	121.2	8.72	5.12	2.36	2.93							
His	H	4	120.7	9.65	5.18	2.73	3.57							
Asn	N	5	116.6	8.37	4.85	2.09	3.07							nd2:113.3 hnd21:6.99 hnd22:7.53
Gln	Q	6	126.6	9.12	4.56	1.93	-	2.64						n2:109.2 hne21:6.50 hne22:7.79
Gln	Q	7	122.2	8.45	-	2.03	2.16	2.57						ne2:112.1 hne21:6.90 hne22:7.58
Ser	S	8	116.3	9.70	3.92	3.99	4.14							
Ser	S	9	119.7	8.31	-	3.96	4.03							
Gln	Q	10	121.2	8.06	4.37	2.03	2.36	2.50	2.63					ne2:111.3 hne21:6.91 hne22:7.41
Pro	P	11		4.59		1.80	2.36	2.17		3.63	3.82			
Pro	P	12												
Thr	T	13	114	8.00	4.79	-								
Thr	T	14	112.9	8.59	5.08	3.81		0.51						
Lys	K	15	118.6	8.81	4.66	1.61	1.77	1.29	-	1.69	-	3.00		
Thr	T	16	118.2	8.52	4.52	4.02		1.14						
Cys	C	17	124.7	8.73	5.09	2.93	3.49							
Pro	P	18			4.35	2.01	2.29	2.13	-	3.64	3.78			
Gly	G	19	111.5	8.80	3.83	4.05								
Glu	E	20	119.1	7.36	4.63	1.93	2.23	2.30	2.41					
Thr	T	21	112	8.78	4.41	-		1.21						
Asn	N	22	118.9	8.24	5.59	2.30	2.64							nd:111.8 hnd16.76 hnd2:6.90
Cys	C	23	116.6	9.07	5.56	2.75	3.30							
Tyr	Y	24	114.2	8.70	6.14	2.86	3.50							hd:6.73 he:6.56
Lys	K	25	125	8.94	5.10	1.63	1.89	1.25	-	1.15	1.57	-	-	
Lys	K	26	131	10.02	5.50	1.86	2.1	1.40	1.68	1.51	-	2.39	2.47	
Val	V	27	124.6	9.15	5.33	2.02		0.93	0.98					
Trp	W	28	128.6	9.16	5.10	3.44	3.49							ne1:130 hne1:10.27 hd1:6.97 hh2:7.18 hz3:6.87 hz2:7.40 he3:7.33 ne:117 hne:7.08
Arg	R	29	118.9	8.39	5.05	1.67	1.80	1.57	-	3.12				
Asp	D	30	124.9	8.46	4.85	2.79	3.18							
His	H	31	116	8.63	4.51									
Arg	R	32	118.6	8.60	4.21	1.82	-	1.17	1.53	3.10	3.18			ne:116.7 hne:7.43
Gly	G	33	108.3	7.89	3.79	4.30								
Thr	T	34	117.6	8.30	4.44	3.87		1.03						
Ile	I	35	131.4	9.02	3.88	0.12		0.68	0.83	-				gCH3:0.27
Ile	I	36	121.4	7.36	5.24	1.46		1.06	-	-				gCH3:0.81
Glu	E	37	128.8	9.53	4.96	2.32		2.61						
Arg	R	38	125	8.75	4.88	1.16		1.45	1.59	3.13	-			ne:116.8 hne:7.16
Gly	G	39	104.6	6.22	3.76	3.86								
Cys	C	40	118.9	8.77	5.17	3.18	3.24							
Gly	G	41	111.2	9.00	3.79	4.22								
Ala	A	42	117.8	7.97	4.57	1.07								
Pro	P	43			4.42	1.68	1.93	0.88	-	2.76	2.91			
Thr	T	44	113.2	7.86	4.16	3.99		1.19						
Val	V	45	119.3	8.25	4.33	2.04	-	0.75	0.83					
Lys	K	46	123.2	7.58	4.53	1.52	1.81	1.37	1.41	1.57	1.62	2.89	-	
Pro	P	47			4.34	1.85	2.33	2.11	-	3.56	3.81			
Gly	G	48	111.5	8.65	3.68	4.33								
Ile	I	49	121.1	7.62	4.43	1.92	-	0.90	1.27	0.43				gCH3:0.75
Lys	K	50	127.4	8.54	4.56	1.87	1.97	1.70	-	1.46	-	3.02	-	
Leu	L	51	128.4	8.70	5.22	1.34	1.87	1.57	-	0.89	-			gCH3:0.76
Asn	N	52	124.6	8.97	5.08	2.71	2.78							nd2:111 hnd21:6.85 hnd22:7.39
Ala	A	53	128.2	8.88	5.68	1.37								
Cys	C	54	118.5	9.17	5.32	3.37	3.69							
Thr	T	55	108.8	8.48	5.04	-		1.21						
Thr	T	56	112.1	7.46	-	4.29		1.20						
Asp	D	57	119.2	8.27	-	-	-							
Lys	K	58	114	9.57	3.27	1.88	2.11	1.02	-	0.51	-	2.77	-	
Cys	C	59	113.1	7.82	4.51	3.38	3.82							
Asn	N	60	121.8	9.15	4.76	2.20	2.82							nd2:117.4 hnd21:7.96 hnd22:8.16
Asn	N	61	121.4	7.53	3.92	2.40	2.57							nd2:111.6 hnd21:6.40 hnd22:7.30

## Liste des déplacements chimiques de Tox62pT3

Tox62pT3	n	Hn	Ha1	Ha2	Hb1	Hb2	Hg1	Hg2	Hd1	Hd2	He1	He2	Hcycle
Leu L 1			4,15		1,44		1,43		0,62	0,73			
Glu E 2	128,2	8,71	5,17		1,75		1,94	2,06					
Cys C 3	119,6	8,60	5,03		2,39	2,96							
His H 4	120,5	9,57	5,19		2,72	3,57							hc4:6,63 hc2:8,05
Asn N 5	117	8,40	4,87		2,11	3,05							nd2:113,2 hnd21:6,99 hnd22:7,53
Gln Q 6	126,6	9,14	4,58		1,91	2,63	2						ne2:109,2 hne21:6,5 hne22:7,79
Gln Q 7	122,2	8,45	4,83		2,03	2,18	2,57						ne2:112,1 hne21:6,91 hne22:7,58
Ser S 8	116,3	9,71	3,94		4	4,15							
Ser S 9	119,6	8,32	4,71		4								
Gln Q 10	121,2	8,07	4,36		2,01	2,34	2,5	2,63					ne2:111,3 hne21:6,91 hne22:7,42
Pro P 11			4,61		1,81	2,38	2,17		3,63	3,84			
Pro P 12			3,89		1,94		2,09		3,63	3,84			
Thr T 13	114	8,04	4,78		4,55		1,22						
Thr T 14	113,5	8,63	5,02		3,8		0,57						
Lys K 15	119	8,77	4,68		1,59	1,75	1,24		1,68	1,71	2,99	3,04	
Thr T 16	120,4	8,56	4,54		3,98		1,31						
Cys C 17	125,9	8,78	4,82		2,77	3,69							
Ser S 18	119,7	9,12	4,63		3,85	4,02							
Pro P 19			4,33		1,91	2,35	2,06	2,17	3,71	3,97			
Gly G 20	111,5	8,68	3,72	4,21									
Glu E 21	121,5	7,79	4,38		1,96	2,33	2,43						
Thr T 22	114,5	8,35	4,43		4,41		1,16						
Asn N 23	119,6	8,60	5,54		2,43	2,67							nd2:112,6 hnd21:6,78 hnd22:7,02
Cys C 24	116	9,14	5,76		2,86	3,27							
Tyr Y 25	114,9	8,88	6,12		2,89	3,42							hd:6,75 He:6,55
Lys K 26	125	8,94	5,11		1,64	1,9	1,16	1,24	1,57		2,71	2,78	
Lys K 27	131	10,01	5,49		1,86	2,12	1,42		1,52		2,4	2,5	
Val V 28	124,8	9,16	5,33		2,04		0,93	1					
Trp W 29	128,7	9,19	5,12		3,44	3,51							ne:130 hne:10,28 hd1:6,98 he3:7,35 hz3:6,89 hz2:7,42 hh2:7,19
Trp W 29b													129,4; 10,12; 7,15
Arg R 30	119	8,40	5,05		1,69	1,81	1,57		3,12				ne:117 hne:7,09
Asp D 31	124,9	8,46	4,86		2,8	3,19							
His H 32	116	8,63	4,54		3,28	3,34							hc4:7,26 hc2:8,35
Arg R 33	118,7	8,60	4,22		1,83		1,17	1,53	3,1	3,19			ne:116,8 hne:7,44
Gly G 34	108,3	7,89	3,79	4,31									
Thr T 35	117,6	8,30	4,46		3,89		1,05						
Ile I 36	131,4	9,03	3,9		0,13		0,66	0,83	0,34				Ch3g : 0,27
Ile I 37	121,5	7,39	5,26		1,47		1,08		0,81				Ch3g : 0,81
Glu E 38	128,5	9,53	4,96		2,33		2,61						
Arg R 39	124,8	8,78	4,87		1,16		1,48	1,59	3,13				ne:116,7 hne:7,27
Gly G 40	104,5	6,22	3,8	3,9									
Cys C 41	119,4	8,60	5,35		3,05	3,26							
Gly G 42	110,9	9,23	3,86	4,3									
Ala A 43	118,3	7,97	4,58		1,12								
Pro P 44			4,32		1,56	1,9	0,92		2,81	2,89			
Thr T 45	113,4	7,88	4,16		4		1,2						
Val V 46	119,7	8,26	4,33		2,04		0,77	0,85					
Lys K 47	123,2	7,53	4,56		1,53	1,82	1,38		1,63		2,89		
Pro P 48			4,36		1,87	2,35	2,02	2,13	3,58	3,83			
Gly G 49	111,5	8,67	3,7	4,34									
Ile I 50	121	7,65	4,46		1,94		0,93	1,29	0,45				CH3g : 0,77
Lys K 51	127,3	8,56	4,58		1,87	1,97	1,72		1,47		3,04		
Leu L 52	128,2	8,70	5,24		1,35	1,89	1,56		0,79				
Asn N 53	124,5	8,96	5,1		2,72	2,81							nd2:111 hnd21:6,86 hnd22:7,41
Ala A 54	127,7	8,87	5,69		1,41								
Cys C 55	118,1	9,12	5,31		3,44	3,71							
Thr T 56	109,1	8,57	5,12		4,81		1,23						
Thr T 57	112,4	7,49	4,78		4,31		1,21						
Asp D 58	119	8,21	4,73		2,44	2,5							
Lys K 59	114,1	9,52	3,27		1,85	1,86	0,49		1		2,78		
Cys C 60	113,1	7,85	4,54		3,43	3,81							
Asn N 61	121,8	9,16	4,76		2,23	2,83							nd2:117,1 hnd21:7,98 hnd22:8,14
Asn N 62	121,6	7,59	3,94		2,44	2,6							nd2:111,7 hnd21:6,44 hnd22:7,33

---

## **ANNEXE 6**

---

Article : Les toxines à trois doigts : même structure, même mécanisme de repliement ?

Soumis à l'occasion des rencontres en toxicologie 2008. Editions de la SFET.



## **Les toxines à trois doigts : même structure, même mécanisme de repliement ?**

Grégori GROSS, Pascal DREVET\*

Institut de Biologie et de Technologies de Saclay, SBIGEM/DSV/CEA, Gif-sur-Yvette, France

\* Auteur correspondant ; Tél : +33 (0)1.69.08.31.31 ; Fax : +33 (0)1.69 08.47.12 ; Courriel: [pascal.drevet@cea.fr](mailto:pascal.drevet@cea.fr)

---

### **Résumé**

Les toxines « à trois doigts » forment un groupe de protéines partageant un repliement commun et exprimant leur activité toxique à travers une multitude de modes d'action. Ces toxines font partie de la famille structurale des protéines à trois doigts qui malgré une grande variabilité dans leurs séquences présentent des éléments structuraux très bien conservés. L'étude des mécanismes de repliement oxydant in vitro de deux toxines à trois doigts montre que ces deux protéines, bien qu'adoptant le même type de structure, se replient selon des mécanismes dissemblables. Ce résultat conforte ceux obtenus lors de l'étude des protéines de type BPTI, qui montraient la disparité de leurs mécanismes de repliement. La présence de quelques éléments conservés semble donc être suffisante pour engendrer une topologie, en revanche les chemins pour parvenir à cette topologie dépendent de facteurs encore inconnus.

### **Three-fingered toxins: same fold, same folding pathway ?**

Three-fingered toxins form a group of proteins sharing a common fold and exhibiting distinct toxic activities. These toxins belong to a larger family of three-fingered proteins which display significant sequence variability but share conserved structural elements. The comparison of the oxidative folding pathway of two three-fingered toxins shows that these two proteins that share the same fold follow different folding pathways. These results agree with former studies conducted on proteins of the BPTI family, showing divergent folding pathways. Conserved structural elements seem to dictate a particular topology whereas the folding pathway may depend on other, still unidentified features.

**Keywords:** three fingered toxin, oxidative folding, folding mechanism, toxic function

---

### **Introduction**

Les toxines « à trois doigts » forment un ensemble de protéines répondant à deux critères : on les trouve dans les venins de serpent et elles adoptent une structure en trois boucles riches en feuillets bêta émergeant d'un cœur compact comportant quatre ponts disulfure (Tsetlin, 1999). Cette communication a pour objet de montrer qu'à l'instar d'une autre famille de protéines, celle du BPTI, les protéines à trois doigts font preuve d'une grande homogénéité de structure, tout en employant des voies de repliement radicalement différentes (Chang et al., 2005).

### **Diversité des modes d'action des toxines à trois doigts**

Malgré une très grande homogénéité structurale les toxines à trois doigts présentent une diversité séquentielle suffisante pour leur conférer un large éventail d'activités.

### **Inhibiteurs de récepteurs impliqués dans le fonctionnement du système nerveux**

Certaines neurotoxines à trois doigts bloquent la communication nerveuse en se liant sur les récepteurs à l'acétylcholine (Servent et Ménez, 2001; Nirthanan and Gwee, 2004). Ces toxines sont exclusivement des antagonistes compétitifs de l'acétylcholine. D'autres

neurotoxines ciblent des récepteurs couplés à la protéine G tels que les récepteurs muscariniques ou adrénergiques. Ces dernières peuvent agir soit comme antagonistes compétitifs, soit comme modulateurs allostériques (Ellis, 1997; Christopoulos and Kenakin, 2002).

### **Inhibiteurs de protéines membranaires non neuronales**

La calciseptine et son homologue la FS2 inhibent les canaux calciques de type L, provoquant le relâchement des muscles lisses et inhibant la contraction cardiaque (de Weille et al., 1991; Yasuda et al., 1994).

La cardiotoxine A5 et la dendroaspine se lient aux intégrines selon des modes de liaison qui diffèrent mais qui dans les deux cas conduisent à l'inhibition de l'agrégation plaquettaire (Rahman et al., 2000 ; Wu et al., 2006).

### **Inhibiteurs d'enzymes**

Les fasciculines se lient à l'acétylcholinestérase et l'inhibent de façon non-compétitive en bloquant la voie d'accès du substrat au site catalytique (Karlsson et al., 1984).

Les hemexines A et B, lorsqu'elles s'associent en un complexe AB, inhibent l'activité protéasique du facteur VIIa ce qui empêche la coagulation (Banerjee et al., 2005).

## Cytotoxines

Il a été montré que leur toxicité s'exerce par la dépolarisation des cardiomyocytes ce qui conduit à un arrêt cardiaque de la victime (Kumar et al., 1997). Le mécanisme proposé pour expliquer ces effets implique la liaison de ces toxines à la membrane des cardiomyocytes, puis une oligomérisation conduisant à la formation de pores aboutissant à la fuite des solutés contenus dans le cytoplasme.

## Eléments communs de structure

La famille des toxines à trois doigts regroupe donc des protéines exerçant leur toxicité selon une grande diversité de modes d'action. Ces toxines appartiennent à la famille structurale plus large des protéines à trois doigts qui regroupe les domaines extracellulaires de divers récepteurs comme le récepteur à l'activine de type II (en anglais, Activin receptor type II), le récepteur de la protéine de la morphogenèse osseuse (BMP receptor), le récepteur du facteur de croissance transformant bêta de type II (TGF beta type II receptor), le récepteur de l'activateur du plasminogène de type urokinase (UPAR),

le récepteur ly-6 et le récepteur CD59 (Tsetlin, 1999). La protéine LYNX qui est associée à la membrane (Miwa et al., 2006) ainsi que la protéine soluble SLURP (Adermann et al. 1999) sont également classées dans la famille des protéines trois doigts.

Les structures cristallographiques de trois exemples de protéines à trois doigts, deux toxines de venin de serpent et le domaine extracellulaire du récepteur du facteur de croissance transformant bêta, sont présentées en figure 1. Ce type de structure est composé de trois longues boucles, ou « doigts », maintenues par quatre ponts disulfure rassemblés dans la partie apicale de la molécule. Il apparaît également que la topologie est conservée même si des adaptations dans les boucles sont manifestes. Les trois doigts de la structure forment deux feuillets bêta antiparallèles reliés par deux ponts disulfure croisés dans la séquence. Cette configuration, appelée en anglais «disulfide beta cross» (Harrison and Sternberg, 1996), est observée dans des protéines adoptant un autre type de repliement et ne suffit donc pas à décrire le repliement à trois doigts.



**Figure 1:** Comparaison des structure cristallographiques de la toxine alpha de *Naja nigricollis* (à gauche), de la cardiotoxine III de *Naja naja atra* (au centre) et du domaine extracellulaire du récepteur du facteur de croissance transformant bêta de type II humain (à droite) (codes PDB:1IQ9, 1H0J et 1M9Z, respectivement).

**Figure 1:** Comparison of the crystallographic structures of toxin alpha from *Naja nigricollis* (left), cardiotoxin III from *Naja naja atra* (center) and the extracellular domain of human TGF beta type II receptor (right) (PDB codes: 1IQ9, 1H0J and 1M9Z, respectively).

Dans le but d'obtenir une définition plus claire du repliement à trois doigts, les séquences des protéines dont l'analyse de la structure tridimensionnelle a confirmé l'existence d'un repliement à trois doigts ont été alignées. Cet alignement comprenant 25 séquences de toxines et 11 séquences de récepteurs a montré que les positions des huit cystéines impliquées dans la formation des quatre ponts disulfure et celle d'une asparagine C-terminale sont strictement conservées (Galat et al., 2008). Il est remarquable que ces résidus soient strictement conservés dans des protéines dont la similarité de séquence est très faible (chacune des séquences de récepteurs a une similarité de séquence avec la toxine alpha inférieure à 25 %). Cela indique que ces neuf acides aminés jouent un rôle prépondérant dans la structuration des protéines à trois doigts. Par ailleurs, l'analyse des structures connues de protéines à trois doigts a montré que trois des ponts conservés ainsi que l'asparagine C-terminale ont une position spatiale fixe et

forment un motif retrouvé dans l'ensemble de ces molécules (Galat et al., 2008). En revanche, le quatrième pont, dont le rôle structural est de fermer la troisième boucle, présente une localisation spatiale variable par rapport au motif conservé. Le motif structural représenté par les trois ponts et l'asparagine C-terminale organisés dans une configuration spatiale définie semble constituer les éléments de structuration suffisant pour définir la structure à trois doigts. La question est donc de savoir si la conservation de ces éléments structuraux implique une conservation du mécanisme par lequel ces protéines se replient.

## Mécanismes de repliement des toxines à trois doigts

D'une manière générale, il est intéressant d'étudier le repliement des protéines réticulées par des ponts disulfure, car l'oxydation de leurs cystéines en cystines

génère des intermédiaires métastables dont le réseau de ponts disulfure est incomplet ; il est possible de bloquer et de purifier ces intermédiaires. La caractérisation physicochimique de ces derniers fournit des informations quant aux étapes franchies par la protéine lors de son repliement. Dans la famille des protéines à trois doigts, les mécanismes de repliement de deux toxines ont été étudiées : celui de la toxine alpha de *Naja nigricollis* et celui de la cardiotoxine III de *Naja naja atra* (Ruoppolo et al., 2001 ; Chang et al., 1998 et 2006)

### Repliement des neurotoxines

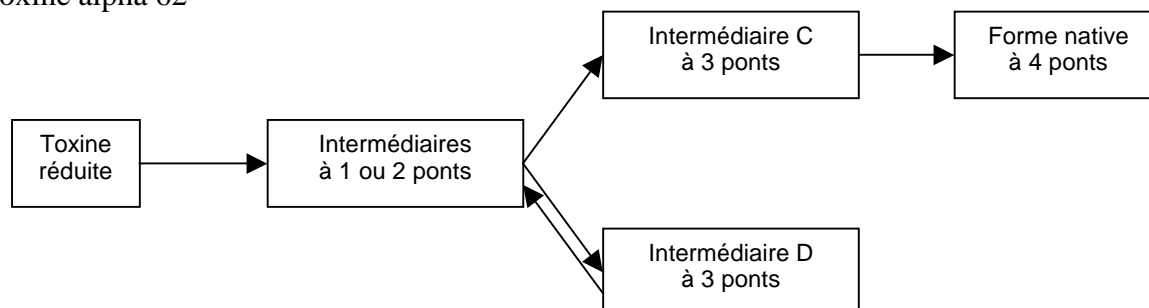
L'étude du repliement oxydant de différentes neurotoxines (Ruoppolo et al., 1998) a montré que la vitesse de repliement dépend de la longueur de la boucle reliant les deux premiers doigts de la molécule (voir figure 1, boucle située au sommet des structures présentées). D'après cette étude, plus cette boucle est petite et plus la vitesse de repliement du variant correspondant est grande. Le mécanisme de repliement d'un mutant de la toxine alpha de *Naja nigricollis* possédant un résidu supplémentaire (Tx alpha 62) a aussi été détaillé (Ruoppolo et al., 2001) : deux intermédiaires à trois ponts natifs mais dont le pont manquant diffère, se forment majoritairement. Ces intermédiaires ont été purifiés, et l'étude de leur repliement a montré qu'un de ces intermédiaires (appelé C) est productif, c'est à dire qu'il peut générer

directement de la protéine native (figure 2). A l'inverse l'autre intermédiaire (appelé D) est une impasse dans le chemin de repliement, car pour générer de la toxine native, il doit d'abord se transformer en intermédiaire C. Enfin, l'étape limitante de la cinétique de repliement de la toxine alpha 62 est la conversion de la forme C vers la forme native. De manière remarquable, le pont manquant de l'intermédiaire C est celui dont la position spatiale est variable au sein de la famille des protéines à trois doigts. Cela suggère que ce pont n'est pas indispensable à l'acquisition de cette structure mais qu'il assume plutôt un rôle stabilisateur.

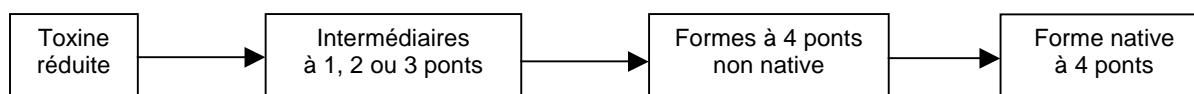
### Repliement de la cardiotoxine III

L'étude du repliement de la cardiotoxine III de *Naja naja atra* a montré que cette toxine se replie selon un mécanisme différent de celui de la toxine alpha (Chang et al., 1998 et 2006). En effet, un massif d'intermédiaires possédant un nombre variable de ponts disulfure apparaît, aucun de ces intermédiaires n'étant majoritaire. De plus, lors de son repliement oxydant, cette protéine produit une quantité importante de formes à quatre ponts, donc complètement oxydées, mais dont l'activité est nulle. D'après le mécanisme proposé par Chang et al. (2006), la conversion de ces intermédiaires à quatre ponts non natifs en la forme native constitue l'étape finale du repliement (voir figure 2).

#### Toxine alpha 62



#### Cardiotoxine III



**Figure 2 :** Comparaison des mécanismes de repliement oxydant de la toxine alpha 62 (en haut) et de la cardiotoxine III (en bas), adapté de Ruoppolo et al. (2001) et Chang et al. (2006).

**Figure 2 :** Comparison of the flow charts of the oxidative folding pathways of toxin alpha 62 (top) and cardiotoxin III (bottom), adapted from Ruoppolo et al. (2001) and Chang et al. (2006).

### Comparaison des deux mécanismes

Deux modèles ont été proposés pour expliquer le mécanisme de repliement des protéines (Daggett and Fersht, 2003). Le modèle « framework » propose que dans les premières étapes, des structures secondaires locales apparaissent puis que dans un deuxième temps ces structures s'associent de manière hiérarchique pour générer la protéine native. A l'inverse dans le modèle « hydrophobic collapse », la première étape est un effondrement désordonné de la molécule sur elle-même

provoqué par des interactions hydrophobes, suivi d'un réarrangement dans un espace confiné.

Dans leur revue sur le repliement des petites protéines à ponts disulfure, Arolas et al. (2006) proposent une classification de ces protéines en fonction du type de mécanisme de repliement. Deux classes principales sont proposées : les protéines se repliant selon un mécanisme de type BPTI d'une part, et les protéines dont le repliement est similaire à celui de l'hirudine d'autre part. D'après ces auteurs la première classe s'apparenterait au modèle « framework » alors



que la seconde serait plus proche du modèle « hydrophobic collapse ». Les résultats obtenus lors de l'étude du repliement oxydant de la toxine alpha 62 et de la cardiotoxine III montrent que la première appartient à la classe de type BPTI (Arolas et al., 2006) et la seconde à la classe de type hirudine (Chang et al., 2006), alors que ces deux protéines partagent des similitudes structurales. Dans le cas de la cardiotoxine III, il est possible que la présence de nombreux résidus hydrophobes nécessaires à la réalisation de la fonction conduise à un repliement de type « hydrophobic collapse ».

## Conclusion

Alors même que les toxines à trois doigts partagent de nombreux éléments de structure, les études du repliement oxydant de deux d'entre elles montrent qu'elles acquièrent leur structure *in vitro* selon des mécanismes différents. Le même type d'analyse a été effectué pour l'inhibiteur de la trypsine pancréatique bovine (BPTI) et le peptide anticoagulant de tique (TAP), qui font partie d'une même famille structurale. Comme pour les toxines à trois doigts, leurs mécanismes de repliement oxydant se sont avérés différents (Chang and Li, 2005). Une disparité dans les mécanismes de repliement au sein d'une même famille structurale ne semble donc pas exceptionnelle.

Cependant, notre connaissance du repliement des protéines reste fragmentaire et repose uniquement sur des expériences menées *in vitro*. Il est possible qu'*in vivo* la situation soit différente. L'influence des protéines chaperons qui agissent sur le repliement, par exemple, reste encore mal définie. En outre, plusieurs études ont suggéré que les protéines commencent à se replier alors que leur synthèse n'est pas terminée (Fedorov and Baldwin, 1997 ; Basharov, 2003). Dans ce cas le repliement de la protéine commence dans la partie N-terminale pendant que la partie C-terminale est encore liée au ribosome, on parle alors de vectorisation du repliement. Des études *in vivo* sont donc indispensables pour apprécier la réalité de la diversité des mécanismes de repliement pour des protéines structuralement proches mais séquentiellement divergentes.

## Références bibliographiques

- Adermann K., Wattler F., Heine G., Meyer M., Forssmann W.G., Nehls M. (1999) Structural and phylogenetic characterization of human SLURP-1, the first secreted mammalian member of the Ly-6/uPAR protein superfamily. *Protein Sci.* 8:810-9.
- Arolas J.L., Aviles F.X., Chang J.-Y., Ventura S. (2006) Folding of small disulfide-rich proteins : clarifying the puzzle. *Trends Biochem. Sci.* 31:292-301.
- Basharov M.A. (2003) Protein folding. *J. Cell. Moll. Med.* 7:223-237.
- Banerjee Y., Mizuguchi J., Iwanaga S., Kini R.M. (2005) Hemextin AB complex, a unique anticoagulant protein complex from hemachatus haemachatus (African Ringhals cobra) venom that inhibits clot initiation and factor VIIa activity. *J. Biol. Chem.* 280:42601-11.
- Chang J.-Y., Kumar T.K.S., Yu C. (1998) Unfolding and refolding of cardiotoxin III elucidated by reversible conversion of the native and scrambled species *Biochemisry* 37:6745-51.
- Chang J.-Y., Li L. (2005) Divergent folding pathways of two homologous proteins, BPTI and tick anticoagulant peptide: compartmentalization of folding intermediates and identification of kinetic traps. *Arch. Biochem. Biophys.* 437:85-95.
- Chang J.-Y., Lu B.-Y., Lin C.C.-J., Yu C. (2006) Fully oxidized scrambled isomers are essential and predominant folding intermediates of cardiotoxin-III. *FEBS Lett.* 580:656-60.
- Christopoulos A., Kenakin T. (2002) G protein-coupled receptor allostery and complexing. *Pharmacol. Rev.* 54:323-74.
- Daggett V., Fersht A.R. (2003) G Is there a unifying mechanism for protein folding ? *Trends Biochem. Sci.* 28:18-25.
- de Weille J.R., Schweitz H., Maes P., Tartar A. Lazdunski M. (1991) Calciseptine, a peptide isolated from black mamba venom, is a specific blocker of the L-type calcium channel. *Proc. Natl. Acad. Sci. USA.* 88:2437-40.
- Ellis J. (1997) Allosteric binding sites on muscarinic receptors. *Drug Dev. Res.* 40:193-204.
- Fedorov A.N., Baldwin T.O. (1997) Cotranslational protein folding *J. Biol. Chem.* 272:32715-18.
- Galat A., Gross G., Drevet P., Sato A., Ménez A. (2008) Conserved structural determinants in three-fingered protein domain. *FEBS J.* 275 :3207-25
- Harrison P.M., Sternberg M.J. (1996) The disulphide beta-cross: from cystine geometry and clustering to classification of small disulphide-rich protein folds. *J. Mol. Biol.* 264:603-23.
- Karlsson E., Mbugua P.M., Rodriguez-Ithurralde D. (1984) Fasciculins, anticholinesterase toxins from the venom of the green mamba *Dendroaspis angusticeps*. *J. Physiol. (Paris)* 79:232-40.
- Kumar T.K., Jayaraman G., Lee C.S., Arunkumar A.I., Sivaraman T., Samuel D., Yu C., (1997) Snake venom cardiotoxins – structure, dynamics, function and folding *J. Biomol. Struct. Dyn.* 15:431-63.
- Miwa J.M., Stevens T.R., King S.L., Caldarone B.J., Ibanez-Tallon I., Xiao C., Fitzsimonds R.M., Pavlides C., Lester H.A., Picciotto M.R., Heintz N. (2006) The prototoxin lynx1 acts on nicotinic acetylcholine receptors to balance neuronal activity and survival *in vivo*. *Neuron* 51:587-600.
- Narayan M., Xu G., Ripoll D.R., Zhai H., Breuker K., Wanjalla C., Leung H.J., Navon A., Welker E., McLafferty F.W., Scheraga H.A. (2004) Dissimilarity in the reductive unfolding pathways of two ribonuclease homologues. *J. Mol. Biol.* 338:795-809.
- Nirathanan S., Gwee M.C.E (2004) Tree-finger aneurotoxins and the nicotinic acetylcholine receptor, forty years on. *J. Pharmacol. Sci.* 94:1-17.
- Rahman S., Flynn G., Aitken A., Patel Y., Hussain F., Lu X., Loftus J.C., French D., Wijelath E., Strand K., Savidge G.F. (2000) Differential recognition of snake venom proteins expressing specific arg-gly-aspartate (RGD) sequence motifs by wild-type and variant integrin  $\alpha$ IIb $\beta$ 3 : further evidence for distinct sites of RGD ligand recognition exhibiting negative allostery. *Biochem. J.* 345:701-9.
- Ruoppolo M., Moutiez M., Mazzeo M.F., Pucci P., Ménez A., Marino G., Quéméneur E. (1998) The length of a single turn controls the overall folding rate of "three-fingered" snake toxins. *Biochemistry* 37:16060-8.
- Ruoppolo M., Talamo F., Pucci P., Moutiez M., Quéméneur E., Ménez A., Marino G. (2001) Slow folding of three-fingered toxins is associated with the accumulation of native disulfide bonded intermediates. *Biochemistry* 40:15257-66.
- Servent D., Ménez A. (2001) Snake neurotoxins that interact with nicotinic acetylcholine receptors. In: Massaro E.J. (Ed.), *Handbook of neurotoxicology*, Humana press inc., Totowa, pp. 385-425.
- Tsetlin V. (1999) Snake venom  $\alpha$ -neurotoxins and other "three-finger" proteins. *Eur. J. Biochem.* 264:281-6.
- Yasuda O., Morimoto S., Jiang B., Kuroda H., Kiruma T., Sakakibara S., Fukuo K., Chen S., Tamatani M., Ogihara T. (1994) FS2 a mamba venom toxin, is a specific blocker of the L-type calcium channels. *Artery.* 21:287-302.



**RESUME :** Le repliement des protéines, dernière étape du processus d'expression de l'information génétique, demeure imparfaitement expliqué. Pendant ma thèse, j'ai étudié le repliement oxydant d'une petite protéine riche en ponts disulfure, la toxine  $\alpha$  de *Naja nigricollis*. J'ai pu montrer que l'addition d'un acide aminé dans une boucle de la structure entraîne un ralentissement global du repliement *in vitro* causé par une permutation des intermédiaires productifs. L'étude en RMN des intermédiaires de repliement a permis de proposer une explication structurale à cette modification de comportement cinétique. Par ailleurs, pour tester l'importance de l'aspect vectoriel du repliement *in vivo*, j'ai développé une méthode ayant pour but de vectoriser le repliement *in vitro*. J'ai pu montrer qu'il est possible grâce à des anticorps dirigés contre la toxine  $\alpha$  réduite d'inhiber son repliement oxydant, de lever cette inhibition et finalement de modifier sa voie de repliement.

**MOTS-CLES :** Repliement des protéines, toxine trois doigts, pont disulfure, modélisation cinétique, structure R.M.N. d'intermédiaires, repliement vectoriel, anticorps polyclonal.

**TITLE :** Modification of the folding pathway of toxin alpha from *Naja nigricollis*: a small disulfide rich protein.

**ABSTRACT:** Protein folding is the last stage of genetic expression. The mechanism by which proteins acquire their 3D structure remains poorly understood. During my project, I showed that the addition of one residue in one loop of the structure of toxin  $\alpha$  from *Naja nigricollis* slows its oxidative folding process down *in vitro*. This decrease in the folding rate is due to a switch of productive pathway. NMR analysis of folding intermediates enabled me to hypothesize a structural explanation for this kinetic modification. Additionally, in order to test the influence of vectorial folding on the folding of a small disulfide-rich protein *in vitro*, I developed a method using antibodies raised against various parts of the reduced protein. My results showed that one of these antibodies is able to inhibit the oxidative folding of toxin  $\alpha$ . This inhibition is reversible. Most interestingly, the use of this antibody modifies the folding pathway.

**KEY WORDS :** Protein folding, three-fingered toxin, disulfide bond, kinetic modelling, NMR structure of intermediate, vectorial folding, polyclonal antibody.